FeedNetBack D03.04 -Wireless Transmission for Control

T. Oechtering, Alain Kibangou, Johannes Karlsson, Ali A. Zaidi, Amirpasha Shirazinia, Lei Bao, Mikael Skoglund

GRANT AGREEMENT N°223866

Deliverable	D03.04
Nature	Report
Dissemination	Public

D03.04 – Wireless Transmission for Control

Report Preparation Date	31/08/2011 Project month: 36
Authors	Tobias J. Oechtering, KTH Alain Kibangou, INRIA, Johannes Karlsson, KTH, Ali A. Zaidi, KTH, Amirpasha Shirazinia, KTH, Lei Bao, KTH, Mikael Skoglund, KTH
Report Version	V1
Doc ID Code	KTH04_R_D0304_310811_V1
Contract Start Date	01/SEP/2008
Duration	41 months
Project Coordinator :	Carlos CANUDAS DE WIT, INRIA, France

Theme 3:

Information and Communication Technologies

SUMMARY

This report documents the research work carried out for Task 3.3. on wireless transmission for control. The main objective is to study and develop physical layer communication strategies which are adapted to fit a given control application. Accordingly, most of the work deal with reliable wireless transmission strategies under strict delay constraints.

Five different approaches are taken. First, two novel transmission schemes for communication over unknown fading channels using non-orthogonal CDMA transmission for multi-sensor and/or multi-agents systems are proposed and discussed. Second, two novel instantaneous source coding mappings for a distributed source-channel coding problem are proposed, based on the study on the Cramér-Rao lower bound and the impact of the stretch factor. The scheme can be used to reduce transmission rate or mitigate the effects of the channel noise in the case of analog transmission. Next, a low-delay cooperative relaying strategy is proposed with an instantaneous mapping function which compresses all received signals in an orthogonal multiple-access relay channel. The optimization of the function results in non-linear mappings which in general outperform linear mappings. Further, the fundamental trade-off between reliability and delay is studied. To this end, upper bounds on the average error probability for sequential coding for the bidirectional broadcast channel with streaming sources are derived and the existence of a deterministic code is proved. Lastly, causal anytime codes based on unequal error protection using Luby transform codes for transmission over symmetric discrete memoryless channel are proposed and analytical upper-bounds on the end-to-end distortion of two anytime transmission schemes are derived.

Contents

1 Introduction

This document is the final report on the work done for *Task 3.3 – Wireless Transmission for Control* of the *Work package 3 – Control and Communications* of the FeedNetback project. The work in Task 3.3 considers wireless transmission strategies where the channel is modeled closer to its physical realization. The main objective of the work is to propose and investigate physical layer communication strategies which can handle delay and reliability constraints required by control applications.

Classical transmission strategies are usually power and/or bandwidth limited and the resulting transmission delays are not a prime constraint. However, for control applications delay is a severe cause of instability. Therefore, most of the work reported in this document considers the study and design of wireless transmission schemes subject to strict delay constraints. Further, the trade-off between delay and reliability of the wireless transmission is discussed. Some results are more theoretical and provide fundamental bounds on the performance which provide conceptual insights to the proposed methods. Some results are more concrete and provide interesting case studies.

When coding over longer time frames is not possible due to delay constraints, the reliability may be increased by spatial redundancy (from many sensors). Accordingly, we considered multi sensor and multiple access channel problems. The process of obtaining channel knowledge in fading channels requires some measurement. The channel sounding strategy might result in an additional possibly unacceptable delay. Therefore, we proposed and studied robust coding strategies for the communication over unknown fading channels. Another research line considers the joint source-channel coding design of memoryless modulation exploiting spatial redundancy from many sensors described by side-information at the receiver, while the memoryless modulation strategy has the lowest possible delay.

The reliability can be also increased by cooperative communication strategy where a relay supports the wireless transmission. However, classical processing strategies for relay channels add a considerable delay to the overall transmission. Therefore, we proposed and studied a strategy based on an instantaneous mapping at the relay which can satisfy strict delay constraints.

Sequential coding strategies are actually able to describe the fundamental trade-off between reliability and delay. The longer the encoder and decoder can wait, the decisions are based on more information and therefore more reliable. Al-

lowing encoder and decoder additionally to take past channel inputs and outputs into account results in the anytime coding strategy. The study and design of communication schemes under this concept directly meet control constraints (anytime coding is basically motivated by control applications). Due to this we derived fundamental bounds for the bidirectional broadcast channel and developed causal anytime transmission scheme based on the rate-less coding strategy.

Outline of the Report

The main body of the report consists of five parts of work regarding wireless transmission for control applications.

First, in Section 2 we consider wireless communication over unknown fading channels using non-orthogonal CDMA transmission for multi-sensor and/or multi-agents systems. A bilinear coding approach and a polynomial precoding approach with corresponding data detection methods based on tensor calculus are proposed. The bit-error-rate performance of the proposed schemes is discussed based on Monte Carlo simulations.

Second, in Section 3 we study a distributed source-channel coding problem where the receiver of a Gaussian channel has additionally access to a correlated version of an analog source. We provide the Cramér-Rao bound and propose two novel mappings based on sinusoidal waveforms, which do not require any encoding or decoding delay. The end-to-end distortion performance with respect to the channel SNR is discussed based on numerical simulations.

Next, in Section 4 we consider an orthogonal multiple-access relay channel where the relay performs an instantaneous mapping to compress all received signals. The mapping strategy is numerically optimized to maximize the achievable sum rate. Further, we provide some solutions of the fixed-point algorithms as examples. The sum-rate performance of the optimized mappings is compared with respect to a linear mapping and the cut-set upper bound.

Fourth, in Section 5 we extend the anytime coding concept to the bidirectional broadcast channel, which is the simplest network model that benefits from the network coding idea. First a sequential encoding and anytime decoding strategy is introduced based on random coding. Then an upper bound on the average error probability is derived. Lastly, the existence of a deterministic code is proved.

Lastly, in Section 6, we study the design of causal anytime codes for transmission over symmetric discrete memoryless channel. We propose an anytime transmission scheme which is based on unequal error protection using Luby transform codes (UEP-LT) and sequential belief propagation (BP) decoding. We also provide a performance analysis on the proposed scheme. In particular, an upper-bound on the end-to-end distortion of the anytime transmission scheme is derived.

Finally, we provide a collection of conclusions of each study in Section 7.

2 Wireless Communication over Unknown Non-orthogonal Fading Channel

The studies reported in this section concern the data transmission over unknown non-orthogonal fading channel in scenarios involving several agents, sensors or actuators. In most of works, the nodes are organized in a network where each sensor processes its individual measurement and transmits the result over an orthogonal multiple-access channel (MAC) to the sink node. In such channels, collisions and interferences between nodes are avoided so that the main impairment of the communication channel concerns noise. Orthogonal MAC can be obtained using TDMA (Time Division Multiple Access), FDMA (Frequency Division Multiple Access) or CDMA (Code Division Multiple Access) protocols. In the first one, the time is divided into slots allocated to each node. Such a scheme induces a latency that can be crucial for feedback control. In the second one, the bandwidth is divided into sub-band allocated to each user. The bandwidth being limited, scalability is a crucial question in this case. For the third scheme, each node is assigned a signaling waveform (or code) generally assumed to be orthogonal, equi-correlated and perfectly correlated with a perfect synchronization in node transmissions (Wimalajeewa and Jayaweera 2007).

In addition to noise, wireless transmissions are also subject to fading. In most works on wireless transmission for control and estimation, fading is in general ignored or assumed to be known. In a recent work, a distributed estimation scheme including a channel estimation using pilot signals was suggested (Senol and Tepedelenlioglu 2008). The derivations were made for parallel channels in a fusion center based wireless sensor network (WSN). It is necessary to point out that the number of parallel channels is limited by the bandwidth and the number of nodes in the network.

Space-time block processing coding and modulation schemes can also be considered. For scenarios in which there is perfect channel state information (CSI), several linear precoding systems have been proposed (see (Sidiropoulos and Budampati 2002) and references therein). However, in practice the CSI at the transmitter suffers from inaccuracies caused by errors in channel estimation and/or limited, delayed or erroneous feedback (Kwon and Cioffi 2008). The derivation of robust coding methods with few or no knowledge on the transmission channel is then a topic of particular interest.

In a scenario with multiple transmitting node and a single sink node having a single antenna, the propagation scenario can be viewed as a highly underdeter-

mined mixture of sources having more sources than sensors. Systems with one single output sensor have received considerably less attention (see (Fernandes, Comon and Favier 2010) and references therein).

By considering non-orthogonal unknown fading channel, multiple transmitting node and a single sink node, we introduce two kinds of precoding where the CSI is not required. In the first one, data are doubly spread before transmission. The precoding can be viewed as bilinear. The second one is a nonlinear precoding scheme. The proposed nonlinear precoding scheme gives rise to a homogeneous Volterra-like input-output equation whose inputs depend on the coding sequence whereas the kernel depends on the informative symbols and on the channel parameters. Both schemes give rise to multidimensional and multilinear data which can be viewed as tensors. Therefore, we first introduce some definitions concerning tensors and multilinear algebra.

2.1 Generalities on Tensors

The order of a tensor $\mathbb{X}$, viewed as a multiway array, is the number of its ways, also called modes, or equivalently the number of indices that characterize each entry. So, a vector is a first-order tensor whereas a matrix is a second-order one. Tensors of order greater than two are called higher-order tensors. The entries of an Nth-order tensor, or N-way array, $\mathbb{X}$ are denoted by $x_{i_1,\cdots,i_N}$, $i_n = 1, 2, \cdots, I_n$, $n = 1, 2, \cdots, N$. In the sequel, we restrict our study to third-order tensors.

A third-order tensor $\mathbb{X} \in \mathcal{C}^{I_1 \times I_2 \times I_3}$ is of rank one if it can be written as an outer product of three vectors $\mathbf{a}^{(n)} \in \mathcal{C}^{I_n}$, $n = 1, 2, 3$:

$$\mathbb{X} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \mathbf{a}^{(3)}. \tag{1}$$

Eq. (1) can be written elementwise as:

$$x_{i_1,i_2,i_3} = a^{(1)}_{i_1} a^{(2)}_{i_2} a^{(3)}_{i_3}. \tag{2}$$

Any third-order tensor $\mathbb{X}$ can be decomposed as a sum of rank-one tensors:

$$\mathbb{X} = \sum_{r=1}^{R} \mathbf{A}^{(1)}_{.r} \circ \mathbf{A}^{(2)}_{.r} \circ \mathbf{A}^{(3)}_{.r}, \tag{3}$$

where $\mathbf{A}^{(n)}_{.r}$ is the r-th column of the factor matrix $\mathbf{A}^{(n)} \in \mathcal{C}^{I_n \times R}$, $n = 1, 2, 3$, and the positive integer R is the number of rank-one components involved in the decomposition. Using (2), this decomposition, called PARAFAC (PARAllel FACtor

analysis), (Harshman 1970), or CANDECOMP (CANonical DECOMPosition), (Caroll and Chang 1970), can be written in the following scalar form:

$$x_{i_1,i_2,i_3} = \sum_{r=1}^{R} a_{i_1,r}^{(n)} a_{i_2,r}^{(n)} a_{i_3,r}^{(n)}, \tag{4}$$

$a_{i_n,r}^{(n)}$ being the entries of the factor matrix $\mathbf{A}^{(n)}$. Using the Kruskal operator (Kruskal 1977, Kolda and Bader 2009), we can write the tensor $\mathbb{X}$ as:

$$\mathbb{X} = [\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}].$$

In the last decade, several PARAFAC or more generally tensor based signal processing methods have been proposed in the literature devoted to communications (Sidiropoulos, Giannakis and Bro 2000, de Almeida, Favier and Mota 2007, Nion and De Lathauwer 2008, Rajih, Comon and Harshman 2008, Kibangou and Favier 2009a). Most of them make use of the spatial diversity induced by multiple receive antennas.

The main property of PARAFAC concerns its essential uniqueness, which means that each factor matrix can be determined up to column scaling and permutation, i.e. two sets of matrices $\{\mathbf{A}^{(n)}\}_{n=1,2,3}$ and $\{\tilde{\mathbf{A}}^{(n)}\}_{n=1,2,3}$ giving rise to the same tensor $\mathbb{X}$ are linked by the following relations $\tilde{\mathbf{A}}^{(n)} = \mathbf{A}^{(n)}\mathbf{\Pi}\mathbf{\Delta}_n$, with $\mathbf{\Delta}_1\mathbf{\Delta}_2\mathbf{\Delta}_3 = \mathbf{I}_R$, where $\mathbf{\Pi}$ and $\mathbf{\Delta}_n$, $n = 1, 2, 3$, are $R \times R$ permutation and diagonal matrices, respectively. A sufficient condition for such a uniqueness, called Kruskal's condition, was established in (Kruskal 1977, Sidiropoulos, Giannakis and Bro 2000) for a third-order tensor: the PARAFAC decomposition (4) is essentially unique if

$$\sum_{n=1}^{3} k_{\mathbf{A}^{(n)}} \geq 2R + 2, \tag{5}$$

where $k_{\mathbf{A}}$ denotes the Kruskal-rank, also called k-rank, of $\mathbf{A}$, i.e. the greatest integer $k_{\mathbf{A}}$ such that any set of $k_{\mathbf{A}}$ columns of $\mathbf{A}$ is independent. The rank and the Kruskal-rank of $\mathbf{A}$ are linked by the following inequality $k_{\mathbf{A}} \leq rank(\mathbf{A})$. Equality occurs when $\mathbf{A}$ is full column rank.

2.1.1 Tensor unfoldings

By fixing one index of the tensor entries, we can define different slices of the tensor:

$$\mathbf{X}_{i_1..} = \begin{pmatrix} x_{i_1,1,1} & \cdots & x_{i_1,1,I_3} \\ \vdots & \ddots & \vdots \\ x_{i_1,I_2,1} & \cdots & x_{i_1,I_2,I_3} \end{pmatrix} \quad \mathbf{X}_{.i_2.} = \begin{pmatrix} x_{1,i_2,1} & \cdots & x_{I_1,i_2,1} \\ \vdots & \ddots & \vdots \\ x_{1,i_2,I_3} & \cdots & x_{I_1,i_2,I_3} \end{pmatrix}$$

$$\mathbf{X}_{..i_3} = \begin{pmatrix} x_{1,1,i_3} & \cdots & x_{1,I_2,i_3} \\ \vdots & \ddots & \vdots \\ x_{I_1,1,i_3} & \cdots & x_{I_1,I_2,i_3} \end{pmatrix}.$$

By concatenating slices along the same mode, we get the three unfolded matrices $\mathbf{X}_1 \in \mathcal{C}^{I_1 I_2 \times I_3}$, $\mathbf{X}_2 \in \mathcal{C}^{I_2 I_3 \times I_1}$,and $\mathbf{X}_3 \in \mathcal{C}^{I_3 I_1 \times I_2}$ given by:

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{X}_{1..} \\ \vdots \\ \mathbf{X}_{I_1..} \end{pmatrix} \quad \mathbf{X}_2 = \begin{pmatrix} \mathbf{X}_{.1.} \\ \vdots \\ \mathbf{X}_{.I_2.} \end{pmatrix}, \quad \mathbf{X}_3 = \begin{pmatrix} \mathbf{X}_{..1} \\ \vdots \\ \mathbf{X}_{..I_3} \end{pmatrix}.$$

A useful feature of PARAFAC is to provide a simple link between the unfolded matrices of a tensor and its factor matrices. For a third-order tensor, we have: $\mathbf{X}_1 = \left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right) \mathbf{A}^{(3)T}$, $\mathbf{X}_2 = \left(\mathbf{A}^{(2)} \odot \mathbf{A}^{(3)}\right) \mathbf{A}^{(1)T}$, $\mathbf{X}_3 = \left(\mathbf{A}^{(3)} \odot \mathbf{A}^{(1)}\right) \mathbf{A}^{(2)T}$, where $\odot$ denotes the Khatri-Rao product[1].

2.1.2 Fitting the PARAFAC model

For a third-order tensor $\mathbb{X}$, computing its PARAFAC decomposition amounts to estimating its factor matrices $\mathbf{A}^{(n)}$, $n = 1, 2, 3$. Various algorithms can be used for estimating these factor matrices (see (Tomasi and Bro 2006) for a comparison of

[1] For $\mathbf{X} \in \mathbb{C}^{I \times R}$ and $\mathbf{Y} \in \mathbb{C}^{J \times R}$, the Khatri-Rao product, denoted by $\odot$, is defined as follows:

$$\mathbf{X} \odot \mathbf{Y} = \begin{pmatrix} \mathbf{Y} diag(\mathbf{X}_{1.}) \\ \vdots \\ \mathbf{Y} diag(\mathbf{X}_{I.}) \end{pmatrix} \in \mathbb{C}^{IJ \times R} \tag{6}$$

It can also be viewed as a column-wise Kronecker product.

$$\mathbf{X} \odot \mathbf{Y} = \begin{pmatrix} \mathbf{X}_{.1} \otimes \mathbf{Y}_{.1} & \cdots & \mathbf{X}_{.R} \otimes \mathbf{Y}_{.R} \end{pmatrix} \in \mathbb{C}^{IJ \times R}, \tag{7}$$

these methods). The alternating least squares (ALS) algorithm (Harshman 1970, Harshman 1972) is the most used one. It acts by alternately minimizing the cost functions in the LS sense

$$\mathcal{J}_1 = \left\| \mathbf{X}_2 - \left(\mathbf{A}^{(2)} \odot \mathbf{A}^{(3)} \right) \mathbf{A}^{(1)T} \right\|_F^2, \quad (8)$$

$$\mathcal{J}_2 = \left\| \mathbf{X}_3 - \left(\mathbf{A}^{(3)} \odot \mathbf{A}^{(1)} \right) \mathbf{A}^{(2)T} \right\|_F^2, \quad (9)$$

$$\mathcal{J}_3 = \left\| \mathbf{X}_1 - \left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)} \right) \mathbf{A}^{(3)T} \right\|_F^2 \quad (10)$$

given initial approximations of two factor matrices. Obviously, only convergence towards a local minimum is guaranteed and the algorithm performance strongly depends on the initialization.

Note that, recently, a non-iterative solution for computing the PARAFAC decomposition when at least one factor matrix has a Toeplitz structure has been proposed (Kibangou and Favier 2009c).

2.2 Bilinear Coding Approach

2.2.1 System model

Let us consider K nodes transmitting their data, with a single antenna, at the same time, within the same bandwidth, towards their neighbors. The kth node has to transmit a sequence $\{s_{j,k}\}_{j=1,\cdots,J}$ after modulation with two spreading waveforms. The modulation scheme can be viewed as a doubly spreading one, (Wong and Lok 2000), i.e. the baseband signal transmitted by the kth node is given by:

$$x_k(t) = \sum_{j=1}^{J} s_{j,k} f_k(t - jT_s), \quad (11)$$

T_s being the symbol period and $f_k(.)$ the modulating waveform given by:

$$f_k(t) = \sum_{q=1}^{Q} b_{q,k} e_k(t - qT_f), \quad e_k(t) = \sum_{i=1}^{I} c_{i,k} g_k(t - iT_c)$$

$g_k(.)$ being a pulse-shape filter. In the sequel, the spreading sequences $\{b_{q,k}\}$ and $\{c_{i,k}\}$, and the pulse-shape filter are assumed to be strictly local and unknown to the other nodes of the network. Note that the informative sequence to be transmitted contains the N entries of the current local information $\boldsymbol{\theta}^{(k)}$ and some additional entries to be specified latter.

In the noiseless case, the baseband signal $y_l(t)$ received by the lth node is given by:

$$y_l(t) = \sum_{k \in \mathcal{K}_l} \beta_{l,k} x_k(t - \tau_{l,k}), \tag{12}$$

where $\beta_{k,l}$ is the fading factor associated with the link between the kth and the lth sensors and $\tau_{k,l}$ the associated delay that holds propagation delay and asynchronism, (Nion and De Lathauwer 2008). By sampling the received signal at time instant $t = jT_s + qT_f + iT_c$, we get:

$$y^{(l)}_{j,q,i} = \sum_{k \in \mathcal{K}_l} s_{j,k} b_{q,k} h^{(l)}_{i,k} \tag{13}$$

where $h^{(l)}_{i,k} = \beta_{l,k} c_{i,k} g_k(t - jT_s - qTf - iT_c - \tau_{l,k})|_{t=jT_s+qT_f+iT_c}$.

The data in (13) can be viewed as a third-order tensor admitting exactly a PARAFAC model. In a compact form, $\mathbb{Y}^{(l)}$, the third-order tensor with $y^{(l)}_{j,q,i}$, $j = 1, \cdots, J$, $i = 1, \cdots, I$, $q = 1, \cdots, Q$, as entries, can be written as follows:

$$\mathbb{Y}^{(l)} = \sum_{k \in \mathcal{K}_l} \mathbf{S}^{(l)}_{.k} \circ \mathbf{B}^{(l)}_{.k} \circ \mathbf{H}^{(l)}_{.k}, \tag{14}$$

meaning that the tensor is completely characterized by the three loading, or factor, matrices $\mathbf{S}^{(l)} \in \Re^{J \times K_l}$, $\mathbf{B}^{(l)} \in \Re^{Q \times K_l}$, and $\mathbf{H}^{(l)} \in \Re^{I \times K_l}$, $\mathcal{K}_l$ being the set of nodes communicating with the lth node, K_l denoting is cardinality.

2.2.2 Data detection method

Applying the Kruskal's condition (5), the factor matrices are essentially unique if

$$k_{\mathbf{S}^{(l)}} + k_{\mathbf{B}^{(l)}} + k_{\mathbf{H}^{(l)}} \geq 2(K_l + 1). \tag{15}$$

We can note that node-wise independent fading and independent design of the pulse-shape filters imply that $\mathbf{H}^{(l)}$ is full rank and full k-rank with very high probability. That is also the case for the $\mathbf{B}^{(l)}$ owing to independence of the spreading sequence $\{b_{q,k}\}$, not restricted to belong to a finite alphabet. We can therefore rewrite the Kruskal condition as follows:

$$k_{\mathbf{S}^{(l)}} + \min(Q, K_l) + \min(I, K_l) \geq 2(K_l + 1). \tag{16}$$

By setting, $Q \geq K_l$ and $I \geq K_l$, we get:

$$k_{\mathbf{S}^{(l)}} \geq 2, \tag{17}$$

meaning that the columns of $\mathbf{S}^{(l)}$ must be pairwise independent.

For all the nodes of the network, the length of the spreading sequences should be chosen such that

$$Q \geq \max(K_l), \quad I \geq \max(K_l). \tag{18}$$

With these conditions, from the received data, each node can retrieve the informations sent by its neighbors up to columns scaling and permutation. Additional information can be included in $\mathbf{S}^{(l)}$ for removing such ambiguities.

Now, given noisy observations $\tilde{\mathbb{Y}}^{(l)}$, we aim to estimate the three factor matrices $\mathbf{H}^{(l)}$, $\mathbf{B}^{(l)}$, and $\mathbf{S}^{(l)}$ using the ALS algorithm. Let us first define the three unfolded matrices associated with $\mathbb{Y}^{(l)}$:

$$\mathbf{Y}_{1,l} = \left(\mathbf{S}^{(l)} \odot \mathbf{B}^{(l)}\right) \mathbf{H}^{(l)T} \in \Re^{JQ \times I}, \tag{19}$$

$$\mathbf{Y}_{2,l} = \left(\mathbf{B}^{(l)} \odot \mathbf{H}^{(l)}\right) \mathbf{S}^{(l)T} \in \Re^{QI \times J}, \tag{20}$$

$$\mathbf{Y}_{3,l} = \left(\mathbf{H}^{(l)} \odot \mathbf{S}^{(l)}\right) \mathbf{B}^{(l)T} \in \Re^{JQ \times I}. \tag{21}$$

When the noise is modeled as temporally and spatially white Gaussian, the maximum likelihood estimation and the least squares fitting result on:

$$\min_{\mathbf{H}^{(l)},\mathbf{B}^{(l)},\mathbf{S}^{(l)}} \left\| \tilde{\mathbf{Y}}_{1,l} - \left(\mathbf{S}^{(l)} \odot \mathbf{B}^{(l)}\right) \mathbf{H}^{(l)T} \right\|_F^2$$

$\tilde{\mathbf{Y}}_{n,l}$, $n = 1, 2, 3$, being the noisy version of $\mathbf{Y}_{1,l}$. It follows that the conditional least squares update of $\mathbf{H}^{(l)}$ is

$$\mathbf{H}^{(l)} = \left(\tilde{\mathbf{Y}}_{1,l} \left(\mathbf{S}^{(l)} \odot \mathbf{B}^{(l)}\right)^{\dagger}\right)^T.$$

Similarly, the conditional least squares updates of the two other matrices are given by

$$\mathbf{S}^{(l)} = \left(\tilde{\mathbf{Y}}_{2,l} \left(\mathbf{B}^{(l)} \odot \mathbf{H}^{(l)}\right)^{\dagger}\right)^T$$

$$\mathbf{B}^{(l)} = \left(\tilde{\mathbf{Y}}_{3,l} \left(\mathbf{H}^{(l)} \odot \mathbf{S}^{(l)}\right)^{\dagger}\right)^T.$$

As stated by (Sidiropoulos, Giannakis and Bro 2000), the conditional update of any given matrix may either improve or maintain but cannot worsen the current fit. Global monotone convergence to (at least) a local minimum follows directly from this observation. The convergence speed can be accelerated if the node knows

the spreading sequence of its neighbors or by using extended line search methods, (Rajih et al. 2008).

After convergence, the node can recover the data contained in $\mathbf{S}^{(l)}$. A preliminary application of this transmission scheme has been recently carried out in a distributed estimation problem (Kibangou 2010).

2.3 Polynomial Precoding Approach

2.3.1 System model

Let us consider K transmitting nodes at the same time and using the same bandwidth. The output at the receiver is then a superposition of K signal waveforms. For each node, the QM-length symbol stream is first parsed into $M \times 1$ symbol vectors $\mathbf{s}_q^{(k)} = \begin{pmatrix} s_{1,q}^{(k)} & \cdots & s_{M,q}^{(k)} \end{pmatrix}^T$, $q = 1, \cdots, Q$. The nonlinear precoding considered herein is a two-stage one. First, each of the symbol vectors $\mathbf{s}_q^{(k)}$ is linearly precoded by an $N \times M$ matrix $\mathbf{A}$. We get $\mathbf{b}_q^{(k)} = \mathbf{A}\mathbf{s}_q^{(k)}$. Note that the linear coding matrix is the same for all the users. Then, the codewords $\mathbf{c}_q^{(k)}$ to be transmitted are obtained through a nonlinear mapping $f(.)$:

$$\mathbf{c}_q^{(k)} = f(\mathbf{b}_q^{(k)}) = f(\mathbf{A}\mathbf{s}_q^{(k)}). \tag{22}$$

The codewords are modulated by a pulse-shape filter $g_k(t)$ so that the baseband signal $x_{k,q}(t)$ transmitted by the kth node is given by

$$x_{k,q}(t) = \sum_{n=1}^{N} c_{n,q}^{(k)} g_k\left(t - (n-1)\,T\right),$$

T being an appropriately chosen fraction of the symbol period T_s.
We assume that each of the signals $x_{k,q}(t)$, $k = 1, \cdots, K$, is received via a single path characterized by a fading-factor α_k and a delay τ_k that holds propagation delay and asynchronism. In baseband, the received signal $y_q(t)$ is then given by:

$$y_q(t) = \sum_{k=1}^{K} \alpha_k x_{k,q}(t - \tau_k) + w_q(t),$$

$w_q(t)$ denoting the additive noise. The discrete-time baseband equivalent model for the received data is then given by:

$$y_{n,q} = y_q(t)|_{t=(n-1)T} = \sum_{k=1}^{K} h_{k,q}\mathbf{c}_{n,q}^{(k)} + w_{n,q}$$

with $h_{k,q} = \alpha_k g_k \left(t-(n-1)\,T-\tau_k\right)|_{t=(n-1)T}$ assumed to be quasi-static, i.e. constant during the transmission of the qth data block.
In the sequel, we consider that the nonlinear function $f(.)$ involved in the encoding process (22) is a pth degree monomial. Therefore, elementwise the received data can be written as follows:

$$\begin{aligned} y_{n,q} &= \sum_{k=1}^{K} h_{k,q} c_{n,q}^{(k)} + w_{n,q} = \sum_{k=1}^{K} h_{k,q} f\Big(\sum_{m=1}^{M} a_{n,m} s_{m,q}^{(k)}\Big) + w_{n,q} \\ &= \sum_{k=1}^{K} h_{k,q} \left(\sum_{m=1}^{M} a_{n,m} s_{m,q}^{(k)}\right)^{p} + w_{n,q} \\ &= \sum_{k=1}^{K} \sum_{m_1=1}^{M} \cdots \sum_{m_p=1}^{M} h_{k,q} \prod_{j=1}^{p} a_{n,m_j} s_{m_j,q}^{(k)} + w_{n,q}. \end{aligned} \tag{23}$$

The aim of our study is to derive estimators of the data symbols $s_{m,q}^{(k)}$ solely from the received data $y_{n,q}$. We assume that the linear precoding matrix $\mathbf{A}$ is known to the receiver. By defining

$$\beta_{m_1,\cdots,m_p,q} = \sum_{k=1}^{K} h_{k,q} \prod_{j=1}^{p} s_{m_j,q}^{(k)}, \tag{24}$$

we can rewrite (23) as:

$$y_{n,q} = \sum_{m_1=1}^{M} \cdots \sum_{m_p=1}^{M} \beta_{m_1,\cdots,m_p,q} \prod_{j=1}^{p} a_{n,m_j}. \tag{25}$$

We can note from (25) that the received signal is linear in the unknown $\beta_{m_1,\cdots,m_p,q}$ but nonlinear in the coding matrix entries. In fact, in a system theory point-of-view, Eq. (25) can be viewed as the input-output equation of a pth-order homogeneous Volterra model (Schetzen 1980), where $\beta_{m_1,\cdots,m_p,q}$ and $a_{n,m}$ represent respectively the Volterra kernel and the input sequence. Moreover, the structure of the kernel (24) looks like that of a parallel cascade Wiener model (see (Kibangou and Favier 2009b)). Therefore, in the sequel, we derive two-stage receivers. The first step consists in estimating the parameters $\beta_{m_1,\cdots,m_p,q}$ in the least squares sense whereas the second one makes use of the algebraic structure of the estimated parameters.

In the sequel, we restrict our study to the third-order case, $p = 3$.

The parameters $\beta_{m_1,m_2,m_3,q}$ can be viewed as entries of a symmetric tensor. Indeed, for any permutation $\pi(.)$ of the indices (m_1,m_2,m_3), we have $\beta_{p_1,p_2,p_3,q} =$

$\beta_{m_1,m_2,m_3,q}$ with $(p_1,p_2,p_3) = \pi(m_1,m_2,m_3)$. We can then rewrite (25), in the noiseless case, as

$$y_{n,q} = \sum_{m_1=1}^{M} \sum_{m_2=m_1}^{M} \sum_{m_3=m_2}^{M} \tilde{\beta}_{m_1,m_2,m_3,q} \prod_{j=1}^{3} a_{n,m_j}. \tag{26}$$

where

$$\tilde{\beta}_{m_1,m_2,m_3,q} = \begin{cases} \beta_{m_1,m_2,m_3,q} & \text{if} \quad m_1 = m_2 = m_3, \\ 3\beta_{m_1,m_2,m_3,q} & \quad\quad m_1 = m_2 \neq m_3, \\ 3\beta_{m_1,m_2,m_3,q} & \quad\quad m_1 = m_3 \neq m_2, \\ 3\beta_{m_1,m_2,m_3,q} & \quad\quad m_2 = m_3 \neq m_1, \\ 6\beta_{m_1,m_2,m_3,q} & \quad\quad m_1 \neq m_2 \neq m_3. \end{cases}$$

In matrix form, Eq. (26) can be written as follows:

$$\mathbf{y}_q = \begin{pmatrix} y_{1,q} & \cdots & y_{N,q} \end{pmatrix}^T = \mathbf{\Phi}\theta_q, \tag{27}$$

where θ_q is a $\bar{Q} \times 1$ vector containing the parameters $\tilde{\beta}_{m_1,m_2,m_3,q}$ to be estimated, $\mathbf{\Phi}$ is an $N \times \bar{Q}$ matrix defined as $\mathbf{\Phi} = \mathbf{\Psi}_A \mathbf{\Omega}$, with $\mathbf{\Psi}_A \begin{pmatrix} \mathbf{A}_{1.} \otimes \mathbf{A}_{1.} \otimes \mathbf{A}_{1.} \\ \vdots \\ \mathbf{A}_{N.} \otimes \mathbf{A}_{N.} \otimes \mathbf{A}_{N.} \end{pmatrix}$, $\mathbf{\Omega}$ is a $M^3 \times \bar{Q}$ column selection matrix, and $\bar{Q} = (M+2)(M+1)M/6$. The least square solution of (27) is given by:

$$\hat{\theta}_q = \mathbf{\Phi}^{\dagger}\mathbf{y}_q \tag{28}$$

provided $\mathbf{\Phi}$ is full column rank. Therefore, the most important criterion for designing the coding matrix $\mathbf{A}$ is to ensure that $\mathbf{\Phi}$ be full column rank. The design of the encoder is then decoupled from the channel knowledge. However, in order to improve the quality of the estimates in a noisy framework it could be necessary to increase N.

Once the parameters $\tilde{\beta}_{m1,m_2,m_3,q}$ have been estimated, we can deduce $\beta_{m1,m2,m3,q}$. Therefore, we will estimate the informative symbols from the estimated parameters $\beta_{m1,m2,m3,q}$, which can be viewed as the entries of a third-order tensor. In the sequel, we remove the index q since the decoding process is per-block.

2.3.2 Data detection using PARAFAC

Let us denote by $\mathbb{B}$ the $M \times M \times M$ third-order symmetric tensor with β_{m_1,m_2,m_3} as entries. From Eq. (24), we can deduce that $\mathbb{B}$ admits a PARAFAC model

(Harshman 1970) with $\mathbf{S}$ and $\mathbf{S}diag(\mathbf{h})$ as factor matrices. Using the Kruskal operator (Kruskal 1977, Kolda and Bader 2009), we get:

$$\mathbb{B} = [\mathbf{S}, \mathbf{S}, \mathbf{S}diag(\mathbf{h})]$$

with

$$\mathbf{h} = \begin{pmatrix} h_{1,q} & \cdots & h_{K,q} \end{pmatrix}^T$$

and

$$\mathbf{S} = \begin{pmatrix} s_{1,q}^{(1)} & \cdots & s_{1,q}^{(K)} \\ \vdots & \ddots & \vdots \\ s_{M,q}^{(1)} & \cdots & s_{M,q}^{(K)} \end{pmatrix}$$

the matrix of the data symbols assumed to be full column rank, which implies $M \geq K$.

From the sufficient condition stated by Kruskal (Kruskal 1977), we can deduce that, the factor matrices are essentially unique, i.e. unique up to column permutation and scaling, if $k_{\mathbf{S}} \geq \frac{2}{3}(K+2)$, where $k_{\mathbf{S}}$ denotes the Kruskal-rank of $\mathbf{S}$. It is also called k-rank and is defined as the greatest integer $k_{\mathbf{S}}$ such that any set of $k_{\mathbf{S}_t}$ columns of $\mathbf{S}$ is independent. Moreover, since the columns of $\mathbf{S}$ are associated with independent users, for $M \geq K$, $\mathbf{S}$ is full column rank with a high probability. As a consequence, the above inequality is always satisfied. Hence, the factor matrices can be obtained up to a scaling factor. The scaling ambiguity can be removed by considering differential modulation or by setting the first row of $\mathbf{S}$ equals to one.

Before deriving the estimation algorithm for fitting the PARAFAC model, we define the following matrix representations of the tensor. The slices of $\mathbb{B}$ are given by

$$\mathbf{B}_{m..} = \mathbf{B}_{.m.} = \mathbf{B}_{..m} = \begin{pmatrix} \beta_{1,1,m} & \cdots & \beta_{1,M,m} \\ \vdots & \ddots & \vdots \\ \beta_{M,1,m} & \cdots & \beta_{M,M,m} \end{pmatrix} = \mathbf{S}diag(\mathbf{S}_{m.})diag(\mathbf{h})\mathbf{S}^T. \tag{29}$$

By concatenating these slices, we get the unfolding matrix

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{1..} \\ \vdots \\ \mathbf{B}_{M..} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_{1..} \\ \vdots \\ \mathbf{B}_{M..} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_{1..} \\ \vdots \\ \mathbf{B}_{M..} \end{pmatrix} = (\mathbf{S} \odot \mathbf{S})\, diag(\mathbf{h}_t)\mathbf{S}^T. \tag{30}$$

For fitting the parameters of the PARAFAC model, we make use of an Alternating least squares algorithm. For this purpose, we define $\mathbf{A}_1 = \mathbf{S}$, $\mathbf{A}_2 = \mathbf{S}$ and $\mathbf{A}_3 = \mathbf{S}diag(\mathbf{h})$, so that we can rewrite the unfolding matrix as follows:

$$\mathbf{B} = (\mathbf{A}_1 \odot \mathbf{A}_2)\,\mathbf{A}_3^T = (\mathbf{A}_2 \odot \mathbf{A}_3)\,\mathbf{A}_1^T = (\mathbf{A}_3 \odot \mathbf{A}_1)\,\mathbf{A}_2^T.$$

The alternating least squares algorithm consists in alternating minimization of the cost functions

$$J_1 = \left\|\mathbf{B} - (\mathbf{A}_2 \odot \mathbf{A}_3)\,\mathbf{A}_1^T\right\|_F^2, \quad J_2 = \left\|\mathbf{B} - (\mathbf{A}_3 \odot \mathbf{A}_1)\,\mathbf{A}_2^T\right\|_F^2,$$
$$J_3 = \left\|\mathbf{B} - (\mathbf{A}_1 \odot \mathbf{A}_2)\,\mathbf{A}_3^T\right\|_F^2.$$

For each cost function, given the two matrices involved in the Khatri-Rao product, the least squares solutions are respectively:

$$\hat{\mathbf{A}}_1^T = (\mathbf{A}_2 \odot \mathbf{A}_3)^\dagger\,\mathbf{B}, \quad \hat{\mathbf{A}}_2^T = (\mathbf{A}_3 \odot \mathbf{A}_1)^\dagger\,\mathbf{B}, \quad \hat{\mathbf{A}}_3^T = (\mathbf{A}_1 \odot \mathbf{A}_2)^\dagger\,\mathbf{B}.$$

After convergence, assuming that $\mathbf{S}$ has 1s as entries of its first row, its estimate is given by:

$$\hat{\mathbf{S}} = \frac{1}{3}\left(\mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3 diag(\mathbf{a})^{-1}\right)$$

with $\mathbf{a}$ as the first row of $\mathbf{A}_3$. We can summarize the ALS estimation method as follows:

1. Initialize $\hat{\mathbf{A}}_1^{(i)}$ and $\hat{\mathbf{A}}_2^{(i)}$, $i = 0$.
2. Increment $i = i + 1$.
3. Compute $\hat{\mathbf{A}}_3^{(i)} = \left(\left(\mathbf{A}_1^{(i-1)} \odot \mathbf{A}_2^{(i-1)}\right)^\dagger \mathbf{B}\right)^T$.
4. Compute $\hat{\mathbf{A}}_1^{(i)} = \left(\left(\mathbf{A}_2^{(i-1)} \odot \mathbf{A}_3^{(i)}\right)^\dagger \mathbf{B}\right)^T$.
5. Compute $\hat{\mathbf{A}}_2^{(i)} = \left(\left(\mathbf{A}_3^{(i)} \odot \mathbf{A}_1^{(i)}\right)^\dagger \mathbf{B}\right)^T$.
6. Go to step 2 until a stoping criterion is reached.
7. Compute $\hat{\mathbf{S}} = \frac{1}{3}\left(\mathbf{A}_1^{(i)} + \mathbf{A}_2^{(i)} + \mathbf{A}_3^{(i)} diag(\mathbf{a}^{(i)})^{-1}\right)$, where $\mathbf{a}^{(i)}$ denotes the first row of $\mathbf{A}_3^{(i)}$.

2.3.3 Data detection based on a matrix joint diagonalization approach

Assuming that $\mathbf{S}$ is full column rank, we can deduce that $\mathbf{S} \odot \mathbf{S}$ is also full column rank (Sidiropoulos, Bro and Giannakis 2000). Obviously, $diag(\mathbf{h})\mathbf{S}^T$ is full row rank, and therefore $rank(\mathbf{B}) = K$, i.e. $\mathbf{B}$ is a rank deficient matrix if $M > K$.

Let us now consider the reduced singular value decomposition (SVD) of $\mathbf{B}$:

$$\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{31}$$

the column-orthonormal matrices $\mathbf{U}$ and $\mathbf{V}$, with respective dimensions $M^2 \times K$ and $M \times K$, containing the left and right singular vectors of $\mathbf{B}$ respectively, whereas the $K \times K$ diagonal matrix $\mathbf{\Sigma}$ is formed with the nonzero singular values of $\mathbf{B}$.

From equations (30) and (31), and the fact that $rank(\mathbf{B}^T) = K$, we deduce that $\mathbf{V}$ and $\mathbf{S}$ span the same column space. So, there exists a nonsingular matrix $\mathbf{F}$, with dimensions $K \times K$, such that

$$\mathbf{S} = \mathbf{V}\mathbf{F}. \tag{32}$$

We can then rewrite the tensor slices as follows:

$$\mathbf{B}_{..m} = \mathbf{V}\mathbf{F}diag(\mathbf{S}_{m.})diag(\mathbf{h})\mathbf{F}^T\mathbf{V}^T.$$

Now, let us define the following symmetric matrices:

$$\mathbf{G}_m = \mathbf{V}^T\mathbf{B}_{..m}\mathbf{V} = \mathbf{F}diag(\mathbf{S}_{m.})diag(\mathbf{h})\mathbf{F}^T, \tag{33}$$

with $m = 1, \cdots, M$. We can conclude that $\mathbf{F}$ jointly diagonalizes the matrices $\mathbf{G}_m$, $m = 1, \cdots, M$. Therefore $\mathbf{F}$ can be obtained by solving a joint diagonalization problem using one of the joint diagonalization algorithms proposed in the literature ((Ziehe, Laskov, Nolte and Müller 2004) for example). Then, $\mathbf{S}$ is estimated using (32). The decoding process is summarized as follows:

1. Compute the matrix $\mathbf{V}$ of the K right singular vectors of $\mathbf{B}$.
2. Construct the set of matrices $\mathbf{G}_m$, $m = 1, \cdots, M$ as follows $\mathbf{G}_m = \mathbf{V}^T\mathbf{B}_{..m}\mathbf{V}$.
3. Find the $K \times K$ matrix $\mathbf{F}$ that jointly diagonalizes the matrices $\mathbf{G}$.
4. Compute the data matrix as $\hat{\mathbf{S}} = \mathbf{V}\mathbf{F}$.

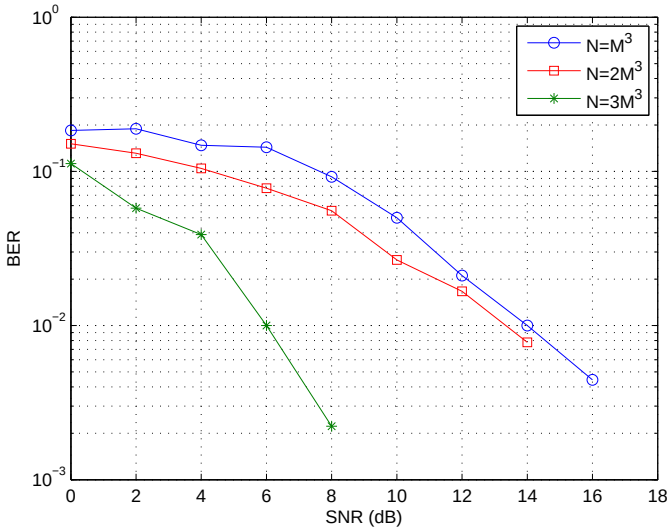


Figure 1: Performance evaluation with different number of rows for the encoding matrix (ALS-PARAFAC case).

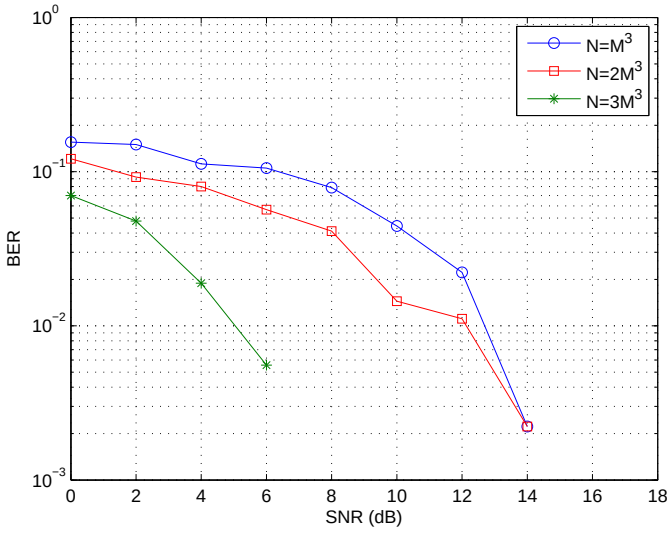


Figure 2: Performance evaluation with different number of rows for the encoding matrix (Joint diagonalization).

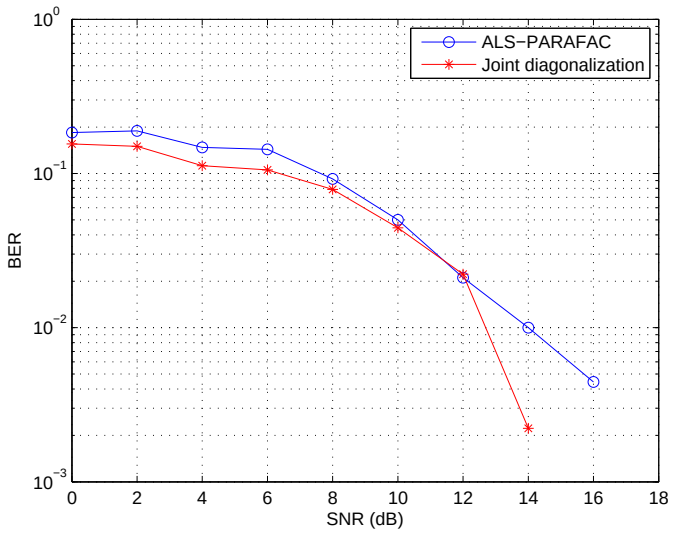


Figure 3: Comparison of the decoding methods ($N = M^3$).

2.3.4 Simulation results

In this section, we give some simulation results. The simulated communication system was characterized by the following parameters: $K = M = 3$. The data sequences was BPSK ones. Both channel parameters and encoding matrix were driven from a uniform distribution. The results presented below are averaged values over 100 Monte Carlo trials. The decoding performance is evaluated in terms of bit-error-rate (BER). The joint diagonalization method used in the second decoding approach is the FFDIAG method (Ziehe et al. 2004).

For each decoding method, in Fig. 1 and 2 we plot the BER according to the signal-to-noise ratio (SNR).

In general, the proposed decoding methods give good results. Significant improvements are obtained by increasing the number N of rows for the encoding matrix $\mathbf{A}$. That is an expected result since by increasing the number of rows for the encoding matrix, the least squares estimation of the data tensor is improved. The improvement is particularly significant for SNR values higher than 2 dB.

In Figures 3, 4, and 5 we compare the two decoding methods for different values of N. We obtain comparable results with both methods. The joint diagonalization approach gives slightly better results. Note that the ALS-PARAFAC

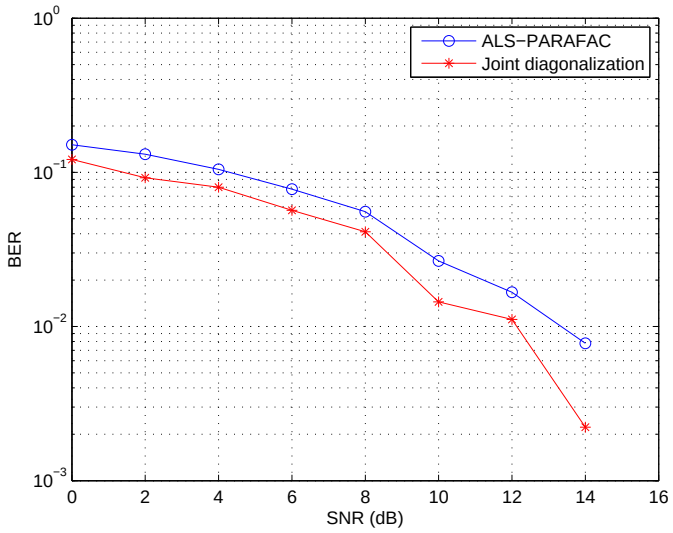


Figure 4: Comparison of the decoding methods ($N = 2M^3$).

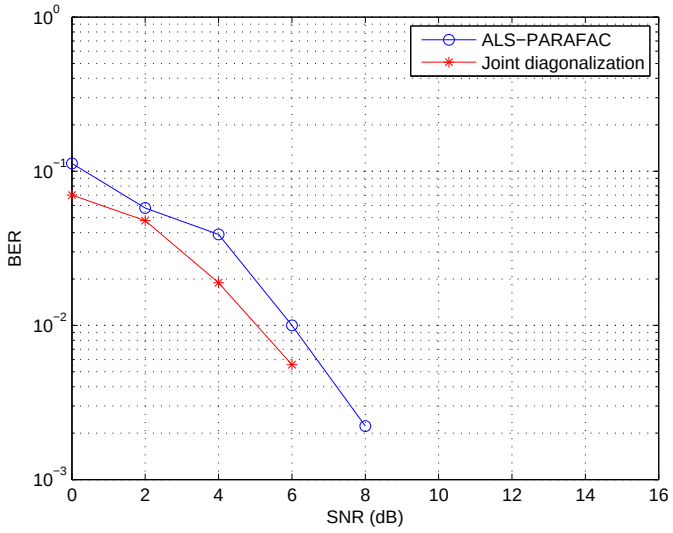


Figure 5: Comparison of the decoding methods ($N = 3M^3$).

were randomly initialized. We considered 10 different initialization and then that giving the best results was selected. The algorithm were stopped after 100 iterations. For these simulations the joint diagonalization approach seems to have more desirable features.

3 Analog Distributed Source–Channel Coding Using Sinusoids

For a controller in a networked control application it is important to have a sufficient precise estimate of its current plant state in order to initiate the correct control action. Assuming the state is measured by a wireless sensor network (observer), we have to find a way to encode and transmit the measurements over the physical channel to a fusion center (receiver), which is connected to the controller. In this work we consider a Gaussian channel without fading. In traditional communication systems source and channel coding are implemented using powerful codes that operate on long blocks of data and, therefore, introduce significant delays. Here, we are interested in transmission subject to strict delay constraints where we cannot use such codes. Therefore, we look at memoryless analog modulation, operating on a sample-by-sample basis, resulting in the lowest possible transmission delay. Cooperative transmission, where spatial correlation structures are exploited, is used to reduce the estimation error and the energy consumption.

3.1 Background

Ever since the work by Wyner and Ziv (Wyner and Ziv 1976), different ways of implementing lossy distributed source coding have been proposed. The most common approach is to do quantization followed by Slepian–Wolf coding (Slepian and Wolf 1973). The Slepian–Wolf code is often implemented using long block codes which introduces significant delays in the decoding process. In situations, such as control applications, where a low delay is of interest, there is much to gain by using a joint source–channel code, where the operations at the sensor node are merged to one single operation — a mapping from the source space directly to the channel space. There are several examples where joint source–channel coding has been used, see (Farvardin and Vaishampayan 1987), (Vaishampayan 1989), (Fuldseth and Ramstad 1997), (Chen and Wornell 1998), (Chung 2000), (Vaishampayan and Costa 2003), (Floor, Ramstad and Wernersson 2007), (Wernersson, Karlsson and Skoglund 2009), (Wernersson, Skoglund and Ramstad 2009), and (Karlsson and Skoglund 2010). In what follows, we propose a novel scheme for *distributed* source–channel coding based on analog mappings, which results in very low transmission delays.

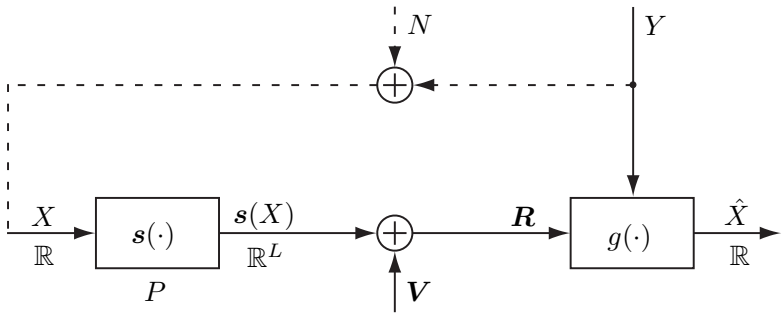


Figure 6: Analog transmission from a sensor node with side information at the receiver. The dashed line shows the structure of the correlation.

3.2 Problem Formulation

Consider a wireless sensor network (WSN), where one of the sensors measures

$$X = Y + N, \tag{34}$$

where Y and N are independent Gaussian random variables with standard deviations σ_Y and σ_N, respectively. The objective is to transmit the sensor's measurement X to a fusion center. Due to spatial correlation between the sensor's measurements, Y is available as side information at the fusion center, see Figure 6. The correlation between X and Y is specified by the correlation coefficient ρ,

$$\rho = \frac{E[XY]}{\sigma_X \sigma_Y} = \frac{\sigma_Y}{\sigma_X}. \tag{35}$$

The channel from the sensor to the fusion center is modeled as a memoryless additive white Gaussian noise (AWGN) channel which can be used L times for each measurement. The sensor encodes X using a mapping $\boldsymbol{s}(X) : \mathbb{R} \mapsto \mathbb{R}^L$. The mapping $\boldsymbol{s}(X)$ should be chosen such that the average power constraint

$$E[\|\boldsymbol{s}(X)\|^2] \leq LP \tag{36}$$

is satisfied. The fusion center receives

$$\boldsymbol{R} = \boldsymbol{s}(X) + \boldsymbol{V}, \tag{37}$$

where $\boldsymbol{V}$ is zero-mean Gaussian noise with covariance matrix $E[\boldsymbol{V}^T\boldsymbol{V}] = \sigma_V^2\boldsymbol{I}$. X is finally estimated from $\boldsymbol{R}$ and the side information Y by a mapping $\hat{X} = g(\boldsymbol{R}, Y) : \mathbb{R}^L \times \mathbb{R} \mapsto \mathbb{R}$. As performance measure we use the mean squared error

$$D = E[(X - \hat{X})^2]. \tag{38}$$

The objective is to find a mapping $\boldsymbol{s}(X)$, satisfying (36), and its corresponding receiver mapping, $g(\boldsymbol{R}, Y)$, such that the distortion D is minimized. In the following, we will discuss general properties of such mappings and evaluate some mappings that are based on sinusoidal waveforms.

3.3 Analysis

The Cramér–Rao lower bound (CRLB) (Trees 1968) gives a lower bound on the distortion for any unbiased receiver mapping. Evaluated for our problem, the CRLB states that

$$D \geq \frac{1}{\frac{1}{\sigma_V^2}E[\|\boldsymbol{s}'(X)\|^2] + \frac{1}{\sigma_N^2}}, \tag{39}$$

where

$$\|\boldsymbol{s}'(X)\| = \sqrt{\sum_{i=1}^{L}\left\{\frac{ds_i(X)}{dX}\right\}^2} \tag{40}$$

is the stretch factor (Wozencraft and Jacobs 1965). From (39), it can be seen how the stretch factor determines how much the channel noise is attenuated; the stretch factor should be as high as possible to limit the effect of the channel noise. One obvious way to accommodate this is by increasing the transmit power but this would violate the power constraint in (36). Another option is to let $\boldsymbol{s}(X)$ be a mapping that twists and bends as much as possible, which means that its derivatives and, consequently, also the stretch factor are high. There is a limit to how much the curve can be stretched by twisting and bending. At some point, different folds of the curve will come too close to each other such that a small channel noise may cause a large decoding error. In conventional source–channel coding, the curve should be stretched as much as possible, keeping different folds of the curve at a maximum distance to avoid large decoding errors (Wozencraft and Jacobs 1965) (Wernersson, Skoglund and Ramstad 2009). In the case of distributed source coding, the side information adds an extra dimension making it possible to have mappings that repeat themselves.

The receiver should form an estimate of x based on the received vector $\boldsymbol{r}$ and the side information y. It is well-known that the optimal estimate, in minimum MSE sense, is given by the conditional expected value

$$\hat{x} = E[X|\boldsymbol{r}, y]. \tag{41}$$

It should be noted that the conditional expected value in general is biased. Nevertheless, the CRLB can still give useful insights into the design of good encoder mappings.

3.4 Proposed Schemes

In this section, we propose two analog distributed source–channel mappings for the case $L = 2$. Our proposed mappings are based on sinusoidal waveforms that thanks to their periodic nature stretch the curve by reusing output symbols. Sinusoids have previously been used for bandwidth expansion of a uniform source with no side information in (Vaishampayan and Costa 2003). In that case the periodic nature of the sinusoids caused some problems and the system had to be designed with a safe margin such that the output symbols does not repeat. However, in our scenario the side information adds an extra dimension making the reuse of output symbols possible.

Mapping 1 is given by

$$\boldsymbol{s}(x) = A \begin{bmatrix} \cos(kx) \\ \sin(kx) \end{bmatrix} \tag{42}$$

$$\|\boldsymbol{s}'(x)\| = Ak \tag{43}$$

where we have also included its stretch factor. The mapping has two parameters, k and A, where k is used to control the periodicity and A is determined by the power constraint in (36). Two examples of this mapping are shown in Figure 7. The mapping in (b), with $k = 5.5$, would be preferable to the one in (a), with $k = 1.8$, due to its higher stretch factor. A direct implication of the higher value of k is that the different turns of the spiral are packed more closely, meaning that the receiver would have to put more trust in the side information. If the correlation is high, this would work well but if the correlation is low, there is a risk for large decoding errors and the mapping in (a) would be more suitable to use.

During simulations we noted that Mapping 1 suffered from performance saturation as the SNR increased. This is because the distortion in the case of a high

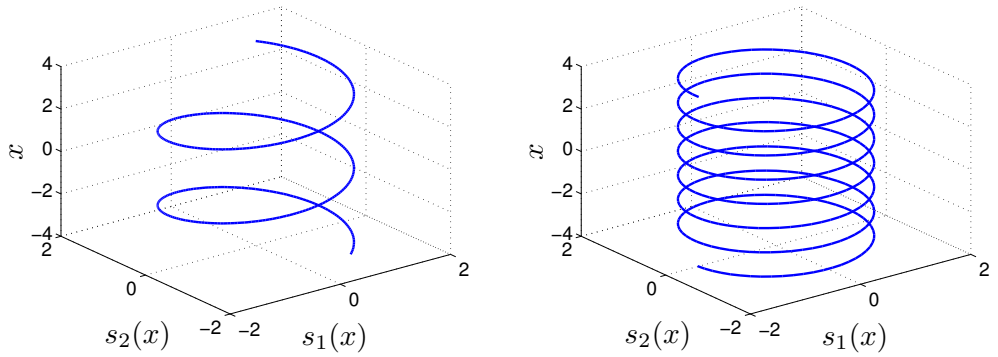


Figure 7: Mapping 1 with, in (a), $k = 1.8$ and, in (b), $k = 5.5$. The figures show how the source symbol x is mapped to a two-dimensional output, $\boldsymbol{s}(x) = [s_1(x)s_2(x)]^T$.

SNR is dominated by errors that occur due to errors in the side information. That is, the transmitted value is estimated as coming from the wrong turn on the spiral. To avoid this behavior, we propose a second mapping, which is a generalization of Mapping 1 with a radius that varies with x.

Mapping 2 is given by

$$\boldsymbol{s}(x) = A(1 + r\cos(mkx)) \begin{bmatrix} \cos(kx) \\ \sin(kx) \end{bmatrix} \tag{44}$$

$$\|\boldsymbol{s}'(x)\| = Ak\sqrt{(1 + r\cos(mkx))^2 + (rm\sin(mkx))^2}. \tag{45}$$

As can be seen, the mapping has two additional parameters, r and m, where r determines how much the radius should vary from unity and m determines the frequency in relation to k at which the radius vary. By letting m be a noninteger, the mapping is repeated less frequently which gives better performance. An example of this mapping can be seen in Figure 8. In this figure $m = 4.5$, meaning that the mapping is repeated every second turn (this is due to the decimal part being .5). One drawback with Mapping 2 is the extra number of parameters, which makes it more complicated to find the optimal combination of parameters.

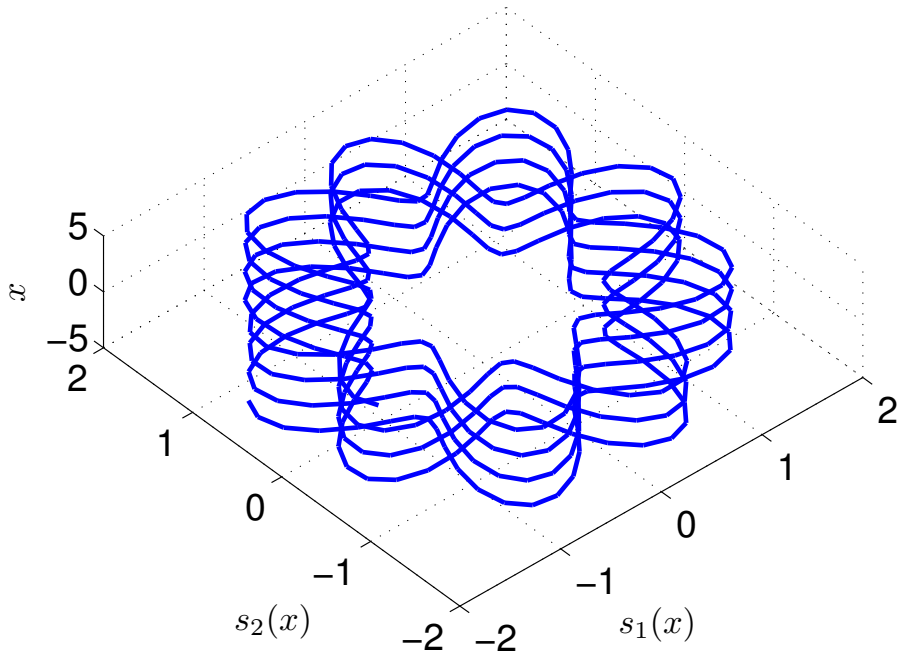
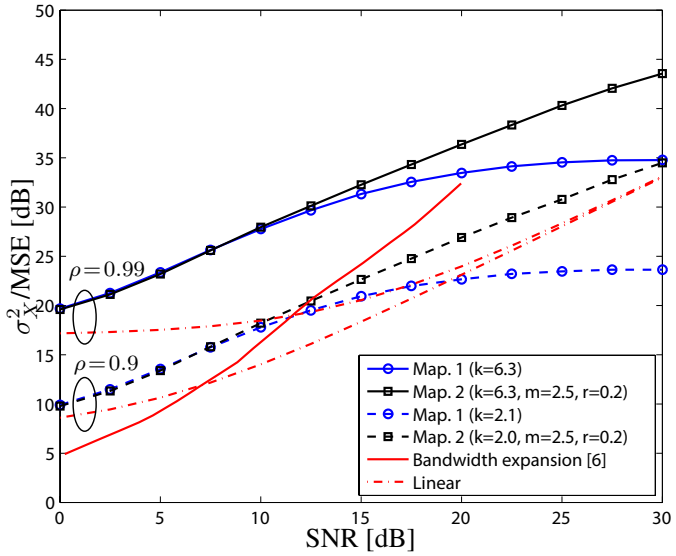


Figure 8: Mapping 2 with $k = 5.6$, $m = 4.5$, $r = 0.35$. The figure shows how the source symbol x is mapped to a two-dimensional output, $\boldsymbol{s}(x) = [s_1(x)s_2(x)]^T$.

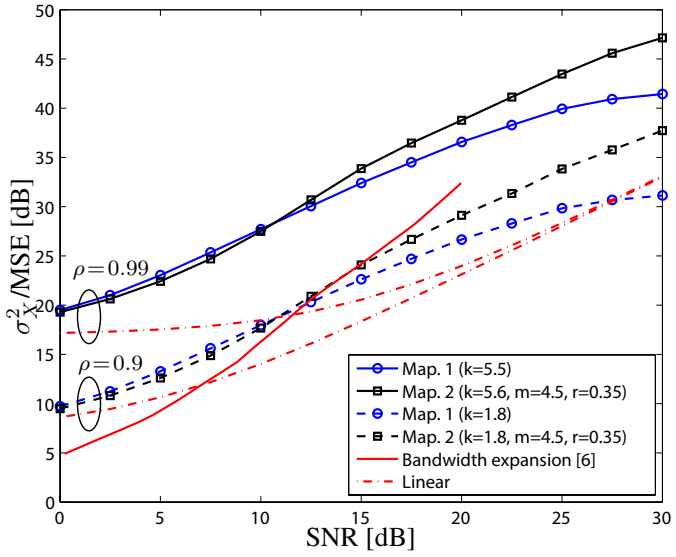
3.5 Numerical Results

The first step in evaluating our proposed mappings is to determine the optimal parameters to use. For Mapping 1 there is only one parameter to determine and we have used a grid search to find good values of k. For Mapping 2 it is a bit more cumbersome to find the optimal parameters since the search space now is three dimensional. In this case we have constrained the search to the sets $m \in \{2.5, 3.5, \ldots, 16.5\}$ and $r \in \{0.2, 0.25, \ldots, 0.45\}$. We consider two different correlation coefficients in our simulations — $\rho = 0.9$ and $\rho = 0.99$. In the case of Mapping 2, we have observed that the parameters m and r seem to only be dependent on the channel SNR and independent of ρ. The performance can be seen in in Figure 9(a) and (b), where the mappings are optimized for an SNR of 5 dB and 15 dB, respectively. We assume that the receiver knows the SNR and therefore adapts to the current channel conditions.

The mappings are evaluated against the 1:2-bandwidth-expansion mappings



(a)



(b)

Figure 9: Performance evaluation. In (a) and (b), the mappings are optimized for SNR=5 dB and SNR=15 dB, respectively.

in (Floor et al. 2007) and also a linear scheme where

$$\boldsymbol{s}(x) = A \begin{bmatrix} x \\ x \end{bmatrix}. \tag{46}$$

The mappings in (Floor et al. 2007) do not utilize any side information but on the other hand they have a higher degree of freedom and have been optimized for each channel-SNR point. They therefore serve as a reference and are used to show the performance gains that come from using side information at the receiver. The linear system uses the optimal receiver, which makes use of the side information, and serves as a reference for the gains that come from a better use and reuse of the available output symbols.

It is expected that the performance improves as the correlation coefficient increases. Interestingly, the proposed mappings manage to keep an approximately constant distance between the curves at $\rho = 0.9$ and $\rho = 0.99$; whereas for the linear system, the advantage of a higher correlation coefficient disappears as the SNR increases. We can also clearly see that Mapping 1 saturates earlier than Mapping 2 when the SNR increases.

3.6 Conclusion

We have proposed the use of analog source–channel mappings, based on sinusoidal waveforms, for implementing distributed source coding. The mappings have been numerically evaluated and are shown to perform well, especially in the case of high correlation and low channel SNR. The proposed transmission scheme requires no encoding or decoding delay, making it suitable for delay-critical control applications. Possible directions for further work is to increase L and also investigate other mappings that can be used.

4 Optimized Analog Network Coding Strategies for the White Gaussian Multiple-Access Relay Channel

In this section we consider a multiple–access relay channel (MARC) with multiple sources, a relay and a destination. We focus on a Gaussian channel model without fading which is closer to the physical transmission. We assume that the source nodes have independent messages to transmit to the destination over orthogonal channels i.e., they either transmit in disjoint frequency bands or disjoint time slots. Furthermore the relay node is working in half-duplex mode, and the received signals at the relay and the destination from different nodes are orthogonal.

The given scenario can be used for wireless transmission for remotely controlling a single vector valued or multiple scalar valued dynamical dynamical systems over a relay channel. Since control applications are delay critical, we assume the relay performs an instantaneous analog mapping to compress all received signals from different sources (state encoders) into one output symbol. In this work, we strived to optimize the memoryless analog relay mapping to maximize the achievable sum rate of the network described under average power constraints of the arbitrarily distributed sources and the relay, which serves as a fundamental bound for further studies.

4.1 Background

The multiple-access relay channel (MARC) consists of multiple source nodes, a single destination node and an intermediate relay node to assist the communication from the multiple sources to the destination. The Gaussian MARC was first introduced by Kramer and Wijngaarden in (Kramer and van Wijingaarden 2000). Capacity inner and outer bounds of the MARC have been studied in (Kramer and van Wijingaarden 2000, Kramer, Gastpar and Gupta 2005, Sankaranarayanan, Kramer and Mandayam 2004b, Sankaranarayanan, Kramer and Mandayam 2004a). The conventional relaying strategies for the classic three-node relay channel (Cover and Gamal 1979), e.g., decode-and-forward (DF) and compress-and-forward (CF) schemes (Kramer et al. 2005, Cover and Gamal 1979, Høst-Madsen and Zhang 2006), can be extended to the MARC. In this work, we specialize our investigation to a MARC scenario where a simple relay with limited processing power is used, c.f. (Gomadam and Jafar 2006, Cui, Ho and Kliewer 2008, Khor-

muji and Larsson 2008, Yao, Khormuji and Skoglund 2008). Essentially the relay is implemented based on an arbitrary function $g : \mathbb{R}^N \to \mathbb{R}$, which maps the N-dimensional input (corresponding to the N sources) to a one-dimensional output. The output is then forwarded to the destination. Compared to the conventional relaying schemes (e.g., DF and CF), the relay scheme considered here is "instantaneous," or "memoryless." The use of such simple relays is motivated by their lower implementation cost and lower processing delay, which is desirable in certain scenarios like large sensor networks.

The mappings considered in this paper "merge" the received signals in the relay before forwarding them to the destination, which resembles the concept of *network coding*, where information is combined in the intermediate nodes (Ahlswede, Cai, Li and Yeung 2000). Hence, we refer to our scheme as *analog network coding*, in the sense that the combining is done in the analog (real-number) domain, in contrast to conventional finite-field network coding requiring decoding before combining. This concept has also been studied in, e.g., (Katti, Maric, Goldsmith, Katabi and Medard 2007, Yang and Koetter 2007, Yao and Skoglund 2009). In this paper, we strive to optimize the mapping g to maximize the achievable sum rate of the network described. Additionally, we introduce an algorithm to obtain an achievable rate region.

4.2 Problem Formulation

We consider the Gaussian multiple-access relay channel with N source nodes $\{\mathcal{S}_i\}_{i=1}^N$, a destination node $\mathcal{D}$ and an intermediate relay node $\mathcal{R}$ as shown in Fig. 10. We assume that the all sources have independent messages to convey to the destination. The encoder at the source $\mathcal{S}_i$ produces the real-valued message symbol X_i with a given probability distribution $f_{X_i}(x_i)$, which is then input to the MARC with an average power P_s. Moreover, all sources transmit their messages on orthogonal channels. Then, $\mathcal{R}$ receives N real-valued inputs and maps them to one real-valued output via a deterministic mapping $g : \mathbb{R}^N \mapsto \mathbb{R}$. As mentioned, we refer to $g(.)$ as an *analog network coding* strategy, as it combines the information coming from all the sources in the network. Furthermore, we assume that the $\mathcal{R} - \mathcal{D}$ link is orthogonal to the $\mathcal{S}_i - \mathcal{D}$ links for all $i \in \{1, 2, ..., N\}$. There are $N + 1$ outputs, given by

$$Y_i = X_i + Z_d^i \quad \forall i \in \{1, 2, ..., N\} \tag{47}$$

$$Y = g(W_1, W_2, ..., W_N) + Z \tag{48}$$

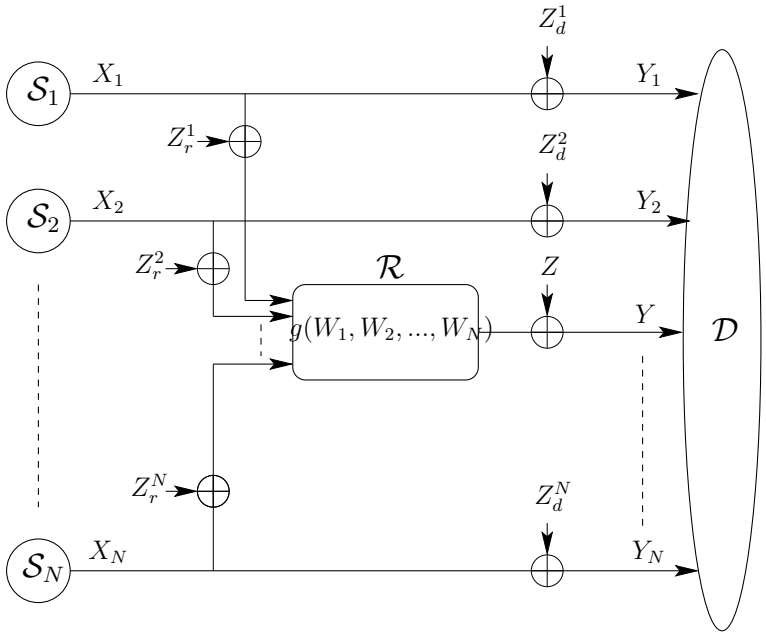


Figure 10: The MARC scenario studied in this paper

where Y_i denotes the received signal from the ith source at the destination and Y denotes the received signal at the destination from the relay. Moreover, $W_i = X_i + Z_r^i$ is the signal received at $\mathcal{R}$ from $\mathcal{S}_i$. We assume that an average power constraint $P_r = E[g^2(W_1, W_2, ..., W_N)]$ is enforced at $\mathcal{R}$. The variables $\{Z_r^i\}_{i=1}^N, \{Z_d^i\}_{i=1}^N$ and Z are mutually independent AWGN components with $Z_r^i \sim \mathcal{N}(0, N_r^i)$, $Z_d^i \sim \mathcal{N}(0, N_d^i)$ and $Z \sim \mathcal{N}(0, N)$. We denote the SNRs of the $\mathcal{S}_i - \mathcal{D}$, $\mathcal{S}_i - \mathcal{R}$ and $\mathcal{R} - \mathcal{D}$ links by $\gamma_{sd}^i, \gamma_{sr}^i$ and γ_{rd} respectively, where $\gamma_{sd}^i = P_s/N_{sd}^i, \gamma_{sr}^i = P_s/N_{sr}^i, \gamma_{rd} = P_r/N$.

Our goal is to optimize the analog network coding strategy at $\mathcal{R}$ to maximize the reliable transmission rate from $\{\mathcal{S}_i\}_{i=1}^N$ to $\mathcal{D}$.

4.2.1 Reference Strategy

One natural choice for the analog mapping $g(.)$ is the linear forwarding strategy. We use linear analog network coding as a benchmark to compare different strategies. The linear strategy is given by

$$g(W_1, W_2, ..., W_N) = \alpha \sum_{i=1}^{N} \beta_i W_i \tag{49}$$

where β_i is the power allocation factor such that $\sum_{i=1}^{N} \beta_i^2 = 1$, and $\alpha = \sqrt{P_r/(P_s + \sum_{i=1}^{N} \beta_i^2 N_{sr}^i)}$ is chosen to satisfy the power constraint at the relay.

4.2.2 Achievable Sum Rate

In the described MARC scenario, the maximum sum-rate at which information can be transmitted reliably from $\{\mathcal{S}\}_{i=1}^{N}$ to $\mathcal{D}$, given the input alphabet distributions $\{f_{X_i}(x_i)\}_{i=1}^{N}$, is obtained as

$$R_{\text{sum}}^{\star} = \max_{\substack{g(W_1,W_2,...,W_N): \\ E[g^2(W_1,W_2,...,W_N)] \leq P_r}} I(X_1, X_2, ..., X_N; Y, Y_1, ..., Y_N) \tag{50}$$

where the rate is measured in bits per channel use [bpcu]. The maximization of mutual information in (50) is done over all possible relay mappings $g : \mathbb{R}^N \mapsto \mathbb{R}$, subject to the relay power constraint. Thus, the relay mapping which maximizes the sum-rate in (50) is given by

$$\begin{aligned} &g^{\star}(W_1, W_2, ..., W_N) = \\ &\arg \max_{\substack{g(W_1,W_2,...,W_N): \\ E[g^2(W_1,W_2,...,W_N)] \leq P_r}} I(X_1, X_2, ..., X_N; Y, Y_1, ..., Y_N) \end{aligned} \tag{51}$$

We refer to $g^{\star}(W_1, W_2, ..., W_N)$ as the *optimal analog network coding strategy*.

Using the chain rule, the mutual information in (50) can be rewritten as

$$\begin{aligned} & I(X_1, X_2, ..., X_N; Y, Y_1, ..., Y_N) \\ = & \sum_{i=1}^{N} I(X_i; Y, Y_1, ..., Y_N | X_1, ..., X_{i-1}) \\ = & \sum_{i=1}^{N} I(X_i; Y, Y_i, ..., Y_N | X_1, ..., X_{i-1}) \\ & + \sum_{i=1}^{N} I(X_i; Y_1, ..., Y_{i-1} | Y, Y_i, ..., Y_N, X_1, ..., X_{i-1}) \\ = & \sum_{i=1}^{N} I(X_i; Y, Y_i, ..., Y_N | X_1, ..., X_{i-1}). \end{aligned} \tag{52}$$

The expansion of mutual information in (52) suggests a sequential decoding scheme at $\mathcal{D}$, where X_1 is decoded first using information from $(Y, Y_1, ..., Y_N)$,

X_2 is decoded next using information from $(Y, Y_2, ..., Y_N, X_1)$, and so on, and the last symbol X_N is decoded using information from $(Y, Y_N, X_1, ...X_{N-1})$. With this sequential decoding at $\mathcal{D}$, the achievable rate of $\mathcal{S}_i$ for any $i \in \{1, 2, ..., N\}$ is given by

$$R_i = I(X_i; Y, Y_i, ..., Y_N | X_1, X_2, ..., X_{i-1}) \tag{53}$$

where the mutual information in (53) is computed using the optimal network coding strategy $g^\star(.)$. Using (50) and (53), the sum-rate can be written as the sum of individual rates

$$R^\star_{\text{sum}} = \sum_{i=1}^{N} R_i. \tag{54}$$

It is noteworthy that $\mathcal{D}$ can decode messages $\{X_i\}_{i=1}^N$ in any order, and thus different rates can be achieved by individual sources for a fixed sum-rate. With N sources, there are $N!$ possible decoding orders at $\mathcal{D}$, leading to $N!$ sets of achievable individual rates $\{R_i\}_{i=1}^N$.

4.2.3 Achievable Rate Region

In order to obtain an achievable rate region, we consider that $\mathcal{R}$ serves only K out of N sources, where $K \in \{1, 2, ..., N\}$. $\mathcal{R}$ selects K sources in any order, then performs a mapping $g : \mathbb{R}^K \mapsto \mathbb{R}$, and drops the remaining $N - K$ inputs. Assuming that $\mathcal{R}$ selects the messages $\{W_i\}_{i=1}^K$ coming from the first K sources $\{\mathcal{S}\}_{i=1}^K$, the maximum achievable sum-rate is then given by

$$R^\star_{\text{sum}} = \sum_{i=K+1}^{N} I(X_i; Y_i) + \max_{\substack{g(W_1, W_2, ..., W_K): \\ E[g^2(W_1, W_2, ..., W_K)] \leq P_r}} I(X_1, X_2, ..., X_K; Y, Y_1, ..., Y_K). \tag{55}$$

For K-source selection at $\mathcal{R}$, there are $\binom{N}{K}$ choices, leading to $\binom{N}{K}$ different sum-rates. For every K-source selection choice, there are $K!$ possible decoding orders, leading to $K!$ sets of individual achievable rates. Thus, for K out of N sources selection, there are $\binom{N}{K}K!$ possible sets of achievable individual rates. The total number of sets of individual achievable rates is then $\sum_{K=1}^{N} \binom{N}{K}K!$, where each set

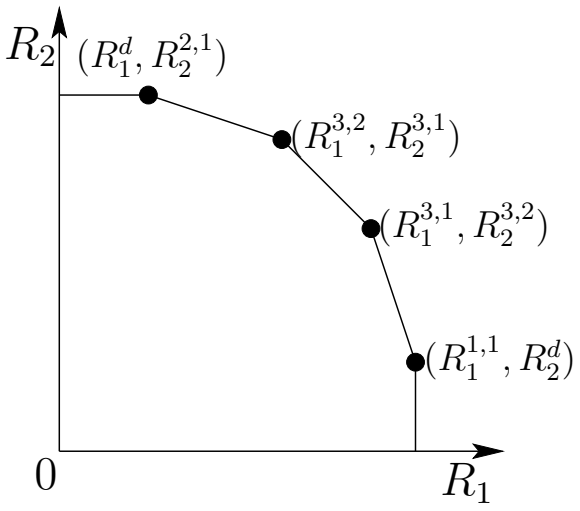


Figure 11: Example: Achievable rate region

represent a point in $\mathbb{R}_+^N$. The achievable rate region is then given by the convex hull of these $\sum_{K=1}^{N} \binom{N}{K} K!$ points.

Example: Consider a MARC with $N = 2$ in which $\mathcal{R}$ can choose to serve any element of $\mathbb{S} = \{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_1\mathcal{S}_2\}$. Let $R_i^{k,j}$ denote the achievable rate of $\mathcal{S}_i$ when $\mathcal{R}$ chooses to serve kth element of set $\mathbb{S}$, and $\mathcal{D}$ chooses to decode the message of $\mathcal{S}_i$ in jth step. Let R_i^d denote the achievable rate of $\mathcal{S}_i$ with direct link only. This leads to four achievable rate points in $\mathbb{R}_+^2$, i.e., $\{(R_1^{1,1}, R_2^d), (R_1^{3,1}, R_2^{3,2}), (R_1^{3,2}, R_2^{3,1}), (R_1^d, R_2^{2,1})\}$. Then, we can obtain an achievable rate region by time-sharing among these points. A typical achievable rate region for two users is illustrated in Fig. 11.

4.3 Numerical Results and Discussion

We devise a method to optimize analog network coding strategies numerically (Zaidi, Khormuji, Yao and Skoglund 2009). This method is based on a fixed point iteration algorithm. In this report we skip the design procedure and only present some numerical examples of optimized analog mappings and the associated achievable rate regions and sum-rate. In the simulations, we consider a symmetric MARC with two sources $\{\mathcal{S}_1, \mathcal{S}_2\}$. The channel is symmetric in the sense that for $\{\mathcal{S}_1, \mathcal{S}_2\}$ the link qualities are the same i.e., $\gamma_{sr}^i = \gamma_{sr}$ and $\gamma_{sd}^i = \gamma_{sd}$ for $i \in \{1, 2\}$. Moreover, we only consider those cases where $\{\mathcal{S}_1, \mathcal{S}_2\}$ have the

same alphabet and probability distribution i.e., $f_{X_i}(x_i) = f_X(x)$ for all $i \in \{1, 2\}$.

4.3.1 Optimized Mappings

Examples of optimized mappings are shown in Figs. 12-15. These figures illustrate how $\mathcal{R}$ should change its strategy based on the quality of $\mathcal{S}_i - \mathcal{R}$ and $\mathcal{S}_i - \mathcal{D}$ links for a fixed quality of $\mathcal{R} - \mathcal{D}$ link. In Figs. 12-14, $\{\mathcal{S}_1, \mathcal{S}_2\}$ have uniform 6-level PAM constellation, and $\mathcal{S}_i - \mathcal{D}$ and $\mathcal{R} - \mathcal{D}$ links have equal strength ($\gamma_{sd} = \gamma_{rd} = 5$ dB). Fig. 12 presents a weak $\mathcal{S}_i - \mathcal{R}$ link scenario ($\gamma_{sr} =$ 5 dB). Interestingly, the optimized $\mathcal{R}$ ignores $\mathcal{S}_2$ and applies the estimate-and-forward (EF) (Gomadam and Jafar 2006) strategy for $\mathcal{S}_1$. This suggests that in order to maximize sum rate, $\mathcal{R}$ should utilize all its power to serve $\mathcal{S}_1$. Fig. 13 presents an optimized mapping for the case with relatively stronger $\mathcal{S}_i - \mathcal{R}$ link ($\gamma_{sr} = 12$ dB), where $\mathcal{R}$ non-linearly maps two inputs to one output. The optimized 2-dimensional mapping is non-invertible and has oscillatory behavior in both dimensions which makes it efficient in utilizing the available power. Using a non-invertible mapping has become possible because the information coming from $\mathcal{R}$ can be decoded at $\mathcal{D}$ using information coming from the direct $\mathcal{S}_i - \mathcal{D}$ links. Fig. 14 shows the optimal mapping for the case with very strong $\mathcal{S}_i - \mathcal{R}$ link ($\gamma_{sr} = 25$ dB). In this case, $\mathcal{R}$ detects the symbol coming from each source and and re-transmits them using a re-arranged constellation(Khormuji and Larsson 2008). We can observe that the re-arranged constellation is also non-invertible since there are 36 possible source outputs and only 8 levels in the re-arranged constellation. In Fig. 15 the optimized mapping has been shown for a higher γ_{sd} (=10 dB) compared to Fig. 13 for the same values of γ_{sr} (=12 dB) and γ_{rd} (=5 dB). Comparison of these two figures indicate that with the increase of γ_{sd}, the oscillatory behavior increases in the optimized mappings because of the more reliable direct $\mathcal{S}_i - \mathcal{D}$ links.

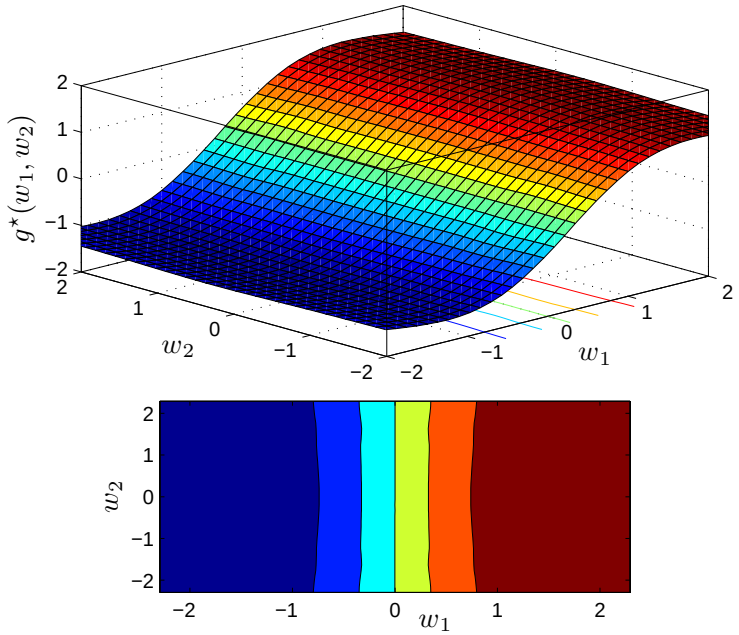


Figure 12: Optimized mapping, γ_{sr}=5 dB, $\gamma_{sd} = \gamma_{rd}$=5 dB

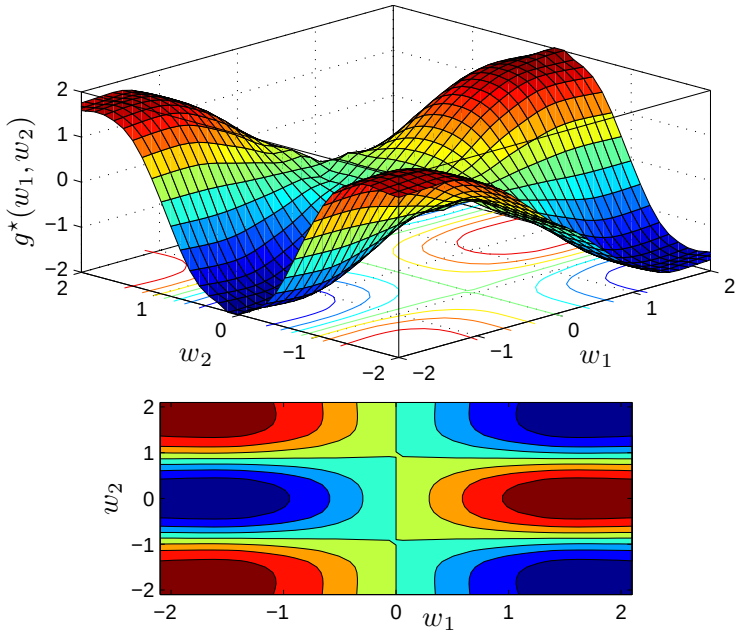


Figure 13: Optimized mapping, γ_{sr}=12 dB, $\gamma_{sd} = \gamma_{rd}$=5 dB

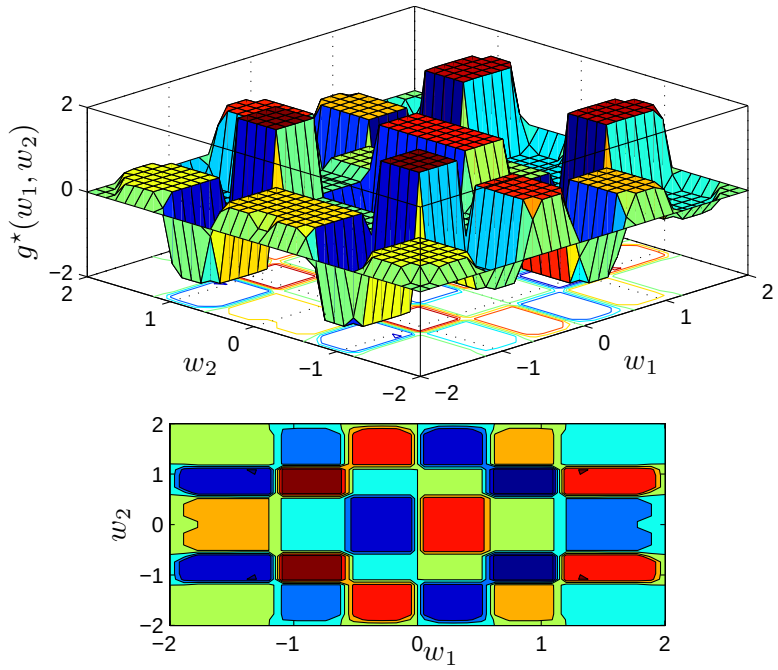


Figure 14: Optimized mapping, γ_{sr}=25 dB, $\gamma_{sd} = \gamma_{rd}$=5 dB

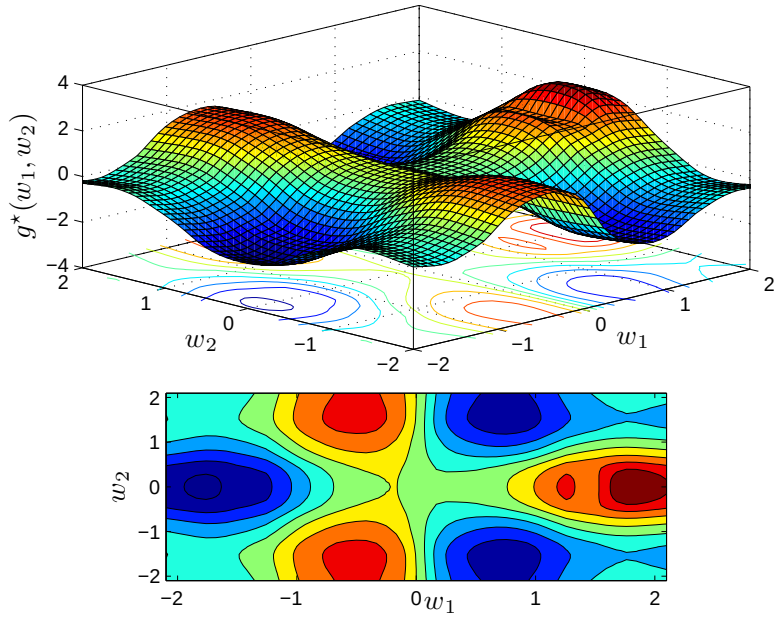


Figure 15: Optimized mapping, γ_{sr}=12 dB, γ_{sd}=10 dB, γ_{rd}=5 dB

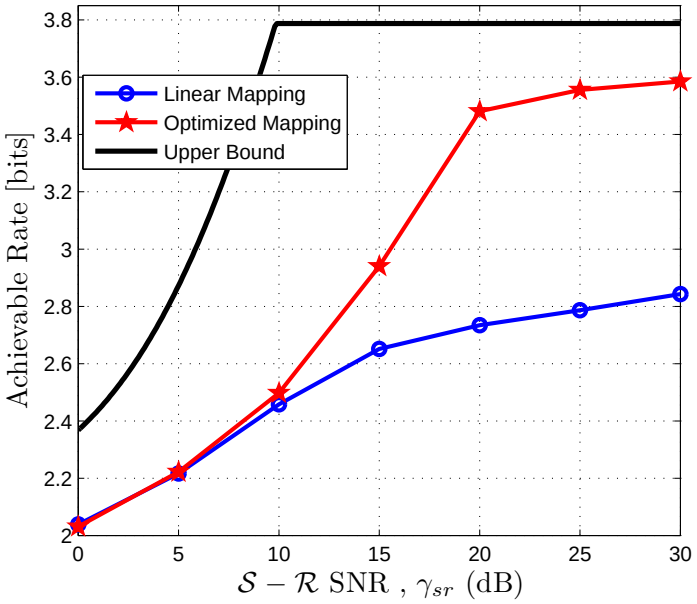


Figure 16: Achievable sum rate. γ_{sd}=5 dB, γ_{rd}=10 dB.

4.3.2 Achievable Rate

In order to evaluate the performance, we numerically compute the achievable sum rate and achievable individual rate for 6-PAM uncorrelated sources using optimized mappings, and compare them against the rates achieved with a linear strategy. Upper bounds to the achievable sum rate and achievable individual rate can be easily derived using the data processing inequality and cut set bound, which are given by

$$\begin{aligned} R_{\text{sum}} \quad \leq \quad & \frac{1}{2}\min\Big\{\sum_{i=1}^{N}\log_2(1+\gamma_{sr}^i+\gamma_{sd}^i), \\ & \log_2(1+\gamma_{rd})+\sum_{i=1}^{N}\log_2(1+\gamma_{sd}^i)\Big\}, \end{aligned} \tag{56}$$

$$\begin{aligned} R_i \quad \leq \quad & \frac{1}{2}\min\Big\{\log_2(1+\gamma_{sr}^i+\gamma_{sd}^i), \\ & \log_2(1+\gamma_{sd}^i)+\log_2(1+\gamma_{rd})\Big\}. \end{aligned} \tag{57}$$

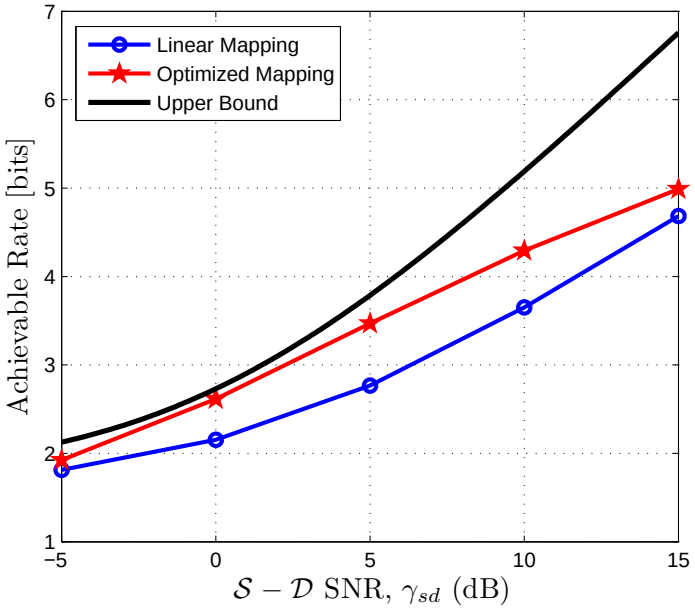


Figure 17: Achievable sum rate. γ_{sr}=30 dB,γ_{rd}=10 dB.

In Fig. 16, the achievable sum-rate has been plotted as a function of γ_{sr} for a fixed γ_{sd} (=5 dB) and γ_{rd} ($=$ 10 dB). We can see from Fig. 16 that the optimized strategy outperforms the linear strategy for all SNRs. As the $\mathcal{S}_i - \mathcal{R}$ links get stronger, the role of $\mathcal{R}$ becomes more significant and hence the gain in the achievable sum-rate with the optimized strategy becomes higher (closer to upper bound). In Fig. 17, the achievable sum-rate has been plotted as a function of γ_{sd} for a fixed γ_s (=30 dB) and γ_{rd} (=10 dB), where also the optimized strategy outperforms the linear strategy for all SNRs. We can observe that for stronger $\mathcal{S}_i - \mathcal{D}$ links, the optimized strategy at $\mathcal{R}$ becomes less useful as more information can be reliably sent with the direct $\mathcal{S}_i - \mathcal{D}$ links. The achievable rate regions with the optimized strategy and linear strategy have been plotted in Fig. 18. To obtain the rate region depicted in Fig. 18, we assume that $\mathcal{R}$ and $\mathcal{D}$ use the relaying and decoding schemes which are explained in section III. The upper bound for rate region is obtained using (56) and (57). Since, the upper bounds were derived assuming Gaussian sources we also plot the achievable rate region with Gaussian source. From Fig. 18, we observe that the optimized strategy significantly enlarges the rate region.

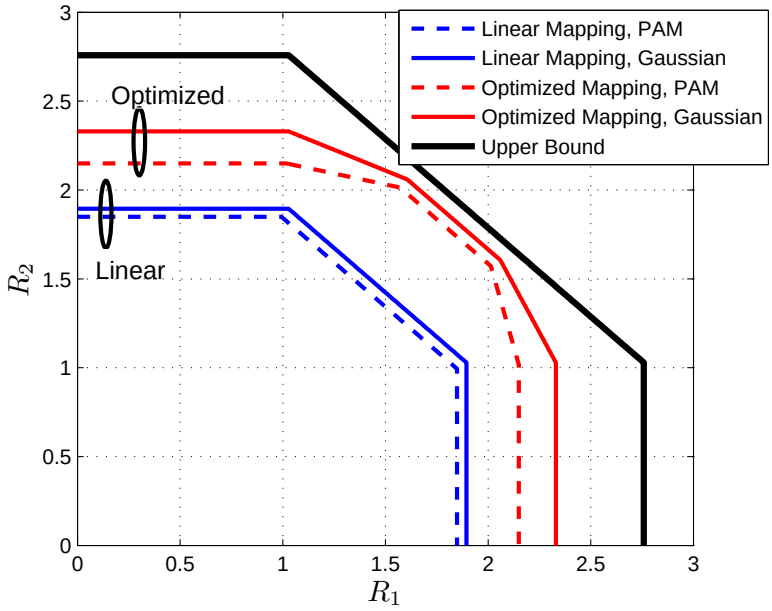


Figure 18: Achievable rate region. γ_{sr}=30 , γ_{sd}=5, and γ_{rd}=10 dB.

4.4 Conclusion

We investigated wireless transmission over Gaussian MARC channel for delay sensitive remote control applications. We provided the design of optimized analog memoryless relay mappings in order to maximize the achievable sum-rate of the network. Our numerical results reveal that the optimized mapping are non-linear in general and they significantly outperform memoryless linear mappings.

5 Coding of Streaming Sources for the Bidirectional Broadcast Channel

Control applications specify a new class of communication problems, where the message to transmit is not necessarily entirely available at the encoder before transmission. If we consider a setup where an observer wants to send the plant state to the control unit via a wireless channel, the information (plant state) arrives at the encoder of the observer successively in time and the controller receives the estimated current state from the decoder in a similar timely manner. Such a setup requires a coding strategy that takes timing evolution into account. Thereby, the decoded information should be as reliable as possible. The degrees of freedom increase when the encoder and decoder can wait for the larger information blocks. However, this results in additional delays which might be unacceptable for the controller and the control objective. The anytime coding concept proposed in (Sahai 2001a, Sahai and Mitter 2006) provides an information-theoretic concept to derive fundamental bounds which characterize the trade-off between reliability and delay.

In this report we extend the anytime coding concept to the bidirectional broadcast channel. The bidirectional broadcast channel is the smallest network where cooperative communication can actually exploit the network coding idea (Ahlswede et al. 2000). Our treatment of the bidirectional channel is motivated from the networked control setup which we have schematically depicted in Figure 19. We consider two separated plants which should be stabilized. The random process at each plant is described by a streaming source. Each controller has access to its own plant state. A joint observer, which has access to both plant states, communicates the other plant state to the controller using a wireless broadcast channel. The motivation of the two plant setup is an assembly line which could be used in an industrial production plant where each production process depends on the other production process.

Outline: In Section 5.1 we provide the information theoretic background to the considered problem, which we formally define in Section 5.2. The main results are in Sections 5.3 and 5.4 where we provide an upper bound to the average error probability which can be achieved using deterministic codes. A conclusion from this work is provided in Section 5.5.

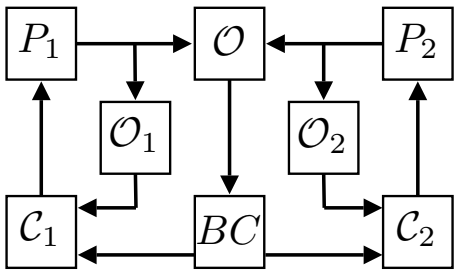


Figure 19: Two plant setup (P_1 and P_2) each with a controller ($\mathcal{C}_1$ and $\mathcal{C}_2$), a joint observer ($\mathcal{O}$), which observes both plant states, and individual observers ($\mathcal{O}_1$ and $\mathcal{O}_2$), which observe one plant state each. The joint observer communicates its observation to the two controllers using a wireless broadcast channel (BC).

5.1 Background

The bidirectional broadcast channel is a two-terminal broadcast channel where the receiving terminals know the message intended for the other terminal. Originally, it describes the second phase of a bidirectional decode-and-forward relaying protocol where two terminals want to communicate with each other using the help of a half-duplex relay node. In the first phase terminals send their messages to the relay node, which decodes the messages. It therefore is described by the classical multiple access channel. In the second phase the relay broadcasts a re-encoded message which allows both terminals to decode the other message using its own source message as side information. Bidirectional relaying using a decode-and-forward strategy based on superposition coding has been introduced in (Rankov and Wittneben 2007). The XOR-coding on the decoded data for bidirectional relaying has been studied first in (Larsson, Johansson and Sunell 2005, Wu, Chou and Kung 2005). The optimal coding strategy combines channel coding for the broadcast channel and the network coding idea. The capacity region of the corresponding bidirectional broadcast channel is independently derived in (Oechtering, Schnurr, Bjelakovic and Boche 2008, Kim, Mitran and Tarokh 2008, Xie 2007).

Theorem 5.1 ((Oechtering et al. 2008))**.** *The capacity region $\mathcal{C}_{BC}$ of the discrete bidirectional memoryless broadcast channel is the convex set of all rate pairs $[R_1, R_2]$ such that*

$$0 \leq R_1 \leq I(X;Y), \quad 0 \leq R_2 \leq I(X;Z),$$

for some input probability mass function $P_X(x)$ and given channel transition probability $P_{YZ|X}(y,z|x)$.

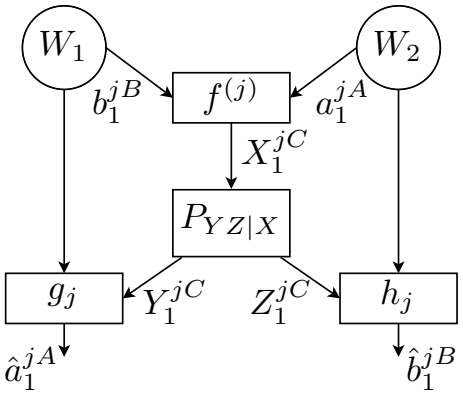


Figure 20: The system diagram of the bidirectional broadcast channel at time instant j with streaming sources W_1 and W_2 which emitted source sequences b_1^{jB} and a_1^{jA} so far. The messages are mapped onto the channel input sequence X_1^{jC} by the encoder $f^{(j)}$. The sequences Y_1^{jC} and Z_1^{jC} denote the channel output of the memoryless broadcast channel $P_{YZ|X}$. Both decoder g_j and h_j make an estimate of the unknown source sequences $\hat{a}_1^{jA}$ and $\hat{b}_1^{jB}$ using the channel output and its side-information. The illustration shows that the information flow in the bidirectional broadcast channel corresponds to the famous butterfly network.

The information flow in a bidirectional relaying protocol corresponds to the famous butterfly network(Ahlswede et al. 2000), which is one of the smallest networks to benefit from network coding. An alternative proof of the coding theorem for the bidirectional broadcast channel has been given in (Oechtering and Skoglund 2010). This proof is based on Gallager's error exponent approach (Gallager 1968, Gallager 1965).

The bidirectional broadcast channel models the communication scenario where each receiver can observe one source and wants to know the other source. In this work we consider streaming sources, which means that both sources steadily emit information while the transmitter sends what it has received so far. This is in contrast to the classical block coding structure where the transmitter knows the whole message before the transmission starts.

Recently, there is an increasing interest in such problems where communication is considered in the context of a control application setup. In (Sahai 2001a, Sahai and Mitter 2006), Sahai and Mitter introduced the concept of anytime reliability for communication of streaming sources, which can be used to derive necessary and sufficient conditions for stabilizing an unstable plant over a noisy com-

munication link. The anytime coding strategy is first extended to multi-terminal problems in (Chang and Sahai 2005b, Chang and Sahai 2005a), where Chang and Sahai study the sequential channel coding for the multiple-access channel and degraded broadcast channel. In (Draper, Chang and Sahai 2005) the ideas are extended to a distributed source coding problem. In this work we extend the sequential coding strategy to the bidirectional communication problem.

5.2 Problem Formulation

We consider sequential encoding and anytime decoding of streaming sources when transmitted over the bidirectional broadcast channel.

5.2.1 Bidirectional Broadcast Channel

We consider a two terminal *discrete memoryless broadcast channel*. The channel input is defined on the finite input alphabet $\mathcal{X}$. The output at terminal 1 and 2 is defined on the finite output alphabets $\mathcal{Y}$ and $\mathcal{Z}$ respectively. Then the broadcast channel is defined by a collection of conditional probability mass function $P_{YZ|X}(y,z|x)$ for all $(x,y,z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. Further, we assume that the channel input does not depend the previous outputs so that the memoryless property implies that for any time instant n we have

$$P(y^n, z^n|x^n) = \prod_{i=1}^{n} P_{YZ|X}(y_i, z_i|x_i),$$

with the notation $x_m^n := (x_m, x_{m+1}, \ldots, x_n)$ with the common convention $x^n = x_1^n$ if $m = 1$. Since the decoder at terminal 1 and 2 do not cooperate, we define the marginal conditional probability mass functions

$$P_1(y^n|x^n) = \prod_{i=1}^{n} \sum_{z_i \in \mathcal{Z}} P_{YZ|X}(y_i, z_i|x_i),$$
$$P_2(z^n|x^n) = \prod_{i=1}^{n} \sum_{y_i \in \mathcal{Y}} P_{YZ|X}(y_i, z_i|x_i).$$

Let W_1 and W_2 denote the random information sources known at the transmitter. In addition, W_1 is known at terminal 1 and W_2 is known at terminal 2. The *bidirectional* communication task for the transmitter is to inform both receiving terminals about their unknown source. In this work we consider streaming

sources, which means that the source sequences are not known before the transmission takes place. Furthermore, both source emit source bits at a different rate than the channel is used. Formally, we assume that there exist integers A, B, and C such that in C channel uses the source W_2 emits a source sequence of length A with letters $a_j \in \{0,1\}$ and source W_1 emits a source sequence of length B with letters $b_j \in \{0,1\}$. Accordingly, the sequential encoding strategy produces channel input sequences of length C. For the following definitions we follow (Chang and Sahai 2005b, Chang and Sahai 2005a).

5.2.2 Sequential Encoding

Let the input alphabet be given by $\mathcal{X} := \{\mathrm{x}_k\}_{k=1}^{|\mathcal{X}|}$. For integers A, B, and C and a probability mass function P_X defined on the input alphabet $\mathcal{X}$, a random sequential encoding scheme is defined by a sequences of encoders $\{f_j\}, j = 1, 2, \ldots$ where,

$$f_j : \{0,1\}^{jA} \times \{0,1\}^{jB} \times [0,1]^C \to \mathcal{X}^C,$$
$$f_j(a_1^{jA}, b_1^{jB}, \lambda_1^C(j, a_1^{jA}, b_1^{jB})) = x_{(j-1)C+1}^{jC}.$$

The random variables $\lambda_i(j, a^{jA}, b^{jB})$, $i = 1, 2, \ldots, C$, $a_1^{jA} \in \{0,1\}^{jA}$, $b_1^{jB} \in \{0,1\}^{jB}$, and $j = 1, 2, \ldots$ provide the common randomness and are independent and uniformly distributed on the interval $[0,1]$. The realizations of the random variables are known at both decoders and encoder and are used to generate a random code. Let $\mathcal{I}_1 := [0, P_X(\mathrm{x}_1)]$ and $\mathcal{I}_k := [\sum_{i=1}^{k-1} P_X(\mathrm{x}_i), \sum_{i=1}^{k} P_X(\mathrm{x}_i)]$, $k = 2, \ldots, |\mathcal{X}|$. such that $\bigcup_{k=1}^{|\mathcal{X}|} \mathcal{I}_k = [0,1]$. Then i-th element of the encoder output sequence f_j for the messages a_1^{jA} and b_1^{jB} is given by

$$x_{(j-1)C+i} = \mathrm{x}_k, \quad \text{if } \lambda_i(j, a^{jA}, b^{jB}) \in \mathcal{I}_k.$$

Further, let $f^{(j)} : \{0,1\}^{jA} \times \{0,1\}^{jB} \to \mathcal{X}^{jC}$ denote the collection of the encoders $f^{(j)} := (f_1, f_2, \ldots, f_j)$, $j = 1, 2, \ldots$. Note that in order to simplify notation, we drop the argument corresponding to common randomness in the definition of $f^{(j)}$. The encoder has the rate pair $R_1 := A/C$ and $R_2 := B/C$ in bits/channel use.

Since for two message sequence pairs (a_1^{nA}, b_1^{nB}) and $(\tilde{a}_1^{mA}, \tilde{b}_1^{mB})$ with $a_1^{kA} = \tilde{a}_1^{kA}$ and $b_1^{kB} = \tilde{b}_1^{kB}$ for some $k \leq \min\{n, m\}$ the first kC encoder outputs are the same, the definition of the sequential encoder is causal and consistent.

5.2.3 Anytime Decoding

For the bidirectional broadcast channel at time jC the decoder of terminal 1 has received the sequence y_1^{jC} and knows the input sequence b_1^{jB}. Similarly, at time jC the decoder of terminal 2 has received the sequence z_1^{jC} and knows the input sequence a_1^{jA}.

The anytime decoding concept means that at any time instant the decoders produce estimates of whole sequences and therefore are allowed to correct past errors. Accordingly, the decoders are defined by sequence $\{g_j\}$ at the receiving terminal 1 with

$$g_j : \{0,1\}^{jB} \times \mathcal{Y}^{jC} \to \{0,1\}^{jA},$$
$$g_j(b_1^{jB}, y_1^{jC}) = \hat{a}_1^{jA}$$

and sequence $\{h_j\}$ at the receiving terminal 2 with

$$h_j : \{0,1\}^{jA} \times \mathcal{Z}^{jC} \to \{0,1\}^{jB},$$
$$h_j(a_1^{jA}, z_1^{jC}) = \hat{b}_1^{jB},$$

$j = 1, 2, \ldots$, where $\hat{a}_1^{jA}$ and $\hat{b}_1^{jB}$ denote the decoded information sequences based on the received channel outputs and side-information so far. Without taking complexity issues into account we consider maximum likelihood decoders at both receivers at any time instant jC, $j = 1, 2, \ldots$. In more detail, if the transmitter has send the codeword $x_1^{jC} = f^{(j)}(a_1^{jA}, b_1^{jB})$ for the source sequences a_1^{jA} and b_1^{jB} and terminal 1 has received the sequence y_1^{jC}, then the decoder g_j decides for the source sequence $\hat{a}_1^{jA}$ if

$$P_1\Big(y_1^{jC} | f^{(j)}(\hat{a}_1^{jA}, b_1^{jB})\Big) \geq \max_{\tilde{a}_1^{jA} \in \{0,1\}^{jA}} P_1\Big(y_1^{jC} | f^{(j)}(\tilde{a}_1^{jA}, b_1^{jB})\Big).$$

Similarly, terminal 2 has received sequence z_1^{jC}, then the decoder h_j decides for the source sequence $\hat{b}_1^{jB}$ if

$$P_2\Big(z_1^{jC} | f^{(j)}(a_1^{jA}, \hat{b}_1^{jB})\Big) \geq \max_{\tilde{b}_1^{jB} \in \{0,1\}^{jB}} P_2\Big(z_1^{jC} | f^{(j)}(a_1^{jA}, \tilde{b}_1^{jB})\Big).$$

If the maximum probability is achieved with multiple source sequences, then the decoders randomly decide between them.

5.3 Upper Bound to Error Probability

For the derivation of the upper bound to the decoding error probability we can follow the derivation of (Chang and Sahai 2005b, Chang and Sahai 2005a). In (Forney 1974) Forney provides an upper bound on the probability of decoding error for the point-to-point channel where the decoder is forced to provide a decision after a certain delay. The result is proved using Gallager's error exponent (Gallager 1968). In (Sahai 2001a) Sahai used this concept in the sequential coding setup.

Theorem 5.2. *Let $\hat{a}_1^{nA}$ and $\hat{b}_1^{nB}$ be the maximum likelihood decoded sequences at terminal 1 and 2 of the source sequences a_1^{nA} and b_1^{nB} after nC channel uses. Then the probability of decoding error of the j-th block can be upper bounded as follows*

$$\mathbb{P}\Big\{\hat{a}_{jA+1}^{(j+1)A} \neq a_{jA+1}^{(j+1)A}\Big\} \leq \mathbb{P}\Big\{\hat{a}_1^{(j+1)A} \neq a_1^{(j+1)A}\Big\} \leq K_1 2^{-dCE_{r,1}(R_1,P_X)}$$

$$\mathbb{P}\Big\{\hat{b}_{jB+1}^{(j+1)B} \neq b_{jB+1}^{(j+1)B}\Big\} \leq \mathbb{P}\Big\{\hat{b}_1^{(j+1)B} \neq b_1^{(j+1)B}\Big\} \leq K_2 2^{-dCE_{r,2}(R_2,P_X)}$$

with decoding delay $dC = (n-j)C$, constant $K_k = (1-2^{-C\,E_{r,k}(R_k,P_X)})^{-1}$, $k = 1, 2$, and Gallager's random error exponents

$$E_{r,k}(R_k, P_X) := \max_{0\leq\rho_k\leq 1} \Big(E_{0,k}(\rho_k, P_X) - \rho_k R_k\Big),\ k = 1, 2$$

$$E_{0,1}(\rho_1, P_X) := -\log_2 \sum_{y\in\mathcal{Y}} \Bigg(\sum_{x\in\mathcal{X}} P_X(x) P_1(y|x)^{\frac{1}{1+\rho_1}} \Bigg)^{1+\rho_1}$$

$$E_{0,2}(\rho_2, P_X) := -\log_2 \sum_{z\in\mathcal{Z}} \Bigg(\sum_{x\in\mathcal{X}} P_X(x) P_2(z|x)^{\frac{1}{1+\rho_2}} \Bigg)^{1+\rho_2}$$

where P_X is the distribution used for the sequential code.

Proof. The proof of the theorem can be deduced from (Forney 1974, Sahai 2001a, Chang and Sahai 2005a) which all follow (Gallager 1965). The details of the proof can be found in (Oechtering and Rathi 2010). □

Remark: From (Oechtering and Skoglund 2010, Gallager 1968) we know that for any rate pair in the interior of the capacity region $\mathcal{C}_{BC}$ we can find a distribution P_X such that the error exponents are positive.

In the next section we show existence of deterministic codes whose error probability is upper bounded by bounds given in Theorem 5.2.

5.4 Existence of Deterministic Codes

In order to show existence of deterministic codes achieving same error performance as that of the codes with common randomness, we extend Sahai's analysis for point-to-point channel to the bidirectional setting (Sahai 2001a). Note that a code with common randomness is equivalent to an ensemble of codes with some probability measure. Thus the performance of a code with common randomness is equal to the average performance of the corresponding ensemble of codes with respect to an appropriate probability measure.

In the following analysis, we will encounter randomness because of the channel and also because of random code selection from the ensemble of codes. We denote the probability measure corresponding to the random code selection by $\mathbf{P}$. As in Section 5.3, we denote average over the ensemble of codes of the probability of an event due to channel randomness by $\mathbb{P}(.)$. For a given encoder $\mathcal{F}$, the probability of an event due to channel randomness is denoted by $\mathbb{P}_{\mathcal{F}}(.)$. In order to further clarify the notation, consider the error event in decoding the j^{th} block at terminal one i.e. $\{\hat{a}_{jA+1}^{(j+1)A} \neq a_{jA+1}^{(j+1)A}\}$. Then,

$$\mathbb{P}\left\{\hat{a}_{jA+1}^{(j+1)A} \neq a_{jA+1}^{(j+1)A}\right\} = E_{\{\mathcal{F}\}}\left(\mathbb{P}_{\mathcal{F}}\left\{\hat{a}_{jA+1}^{(j+1)A} \neq a_{jA+1}^{(j+1)A}\right\}\right).$$

In the following lemma, we show that average error probability over any set of encoders with positive probability has same behavior as that of the average over the whole ensemble of encoders.

Lemma 5.3. *Consider transmission over bidirectional broadcast channel with sequential encoding and anytime decoding as defined in Section 5.2. Let* $\mathbf{P}$ *be the probability measure defined on the set of encoders* $\{\mathcal{F}\}$ *with rate pair* (R_1, R_2)*. The rates are rational numbers* $R_1 = A/C$*,* $R_2 = B/C$*. If the average error probability for decoders 1 and 2 decay exponentially with delay* d *as*

$$\begin{aligned}\mathbb{P}\left\{\hat{a}_{jA+1}^{(j+1)A} \neq a_{jA+1}^{(j+1)A}\right\} &\leq K_1 2^{-dC\alpha_1},\\ \mathbb{P}\left\{\hat{b}_{jB+1}^{(j+1)B} \neq b_{jB+1}^{(j+1)B}\right\} &\leq K_2 2^{-dC\alpha_2},\end{aligned}$$

then the average error probability over any set with strictly positive probability decays with delay with the same exponent.

Proof. The proof is done by contradiction and can be found in (Oechtering and Rathi 2010). □

The previous lemma shows that there can not be a "large" set of encoders whose performance is worse than the average. However it does not say anything about the performance of a deterministic code. The next lemma deals with this issue.

Lemma 5.4. *Consider the set of encoders $\{\mathcal{F}\}$ with the probability measure $\mathbf{P}$ for the bidirectional broadcast channel for the rate pair $R_1 = A/C$ and $R_2 = B/C$. Suppose that the average probability of error is upper bounded as*

$$\mathbb{P}\left\{\hat{a}_{jA+1}^{(j+1)A} \neq a_{jA+1}^{(j+1)A}\right\} \leq K_{1,j}2^{-dC\alpha_1}, \tag{58}$$

$$\mathbb{P}\left\{\hat{b}_{jB+1}^{(j+1)B} \neq b_{jB+1}^{(j+1)B}\right\} \leq K_{2,j}2^{-dC\alpha_2}, \tag{59}$$

for all $d \geq 0$. Then for almost every deterministic encoder $\mathcal{F}$, every $\epsilon > 0$, there are positive constants $K_{1,j}^{\epsilon,\mathcal{F}}$ and $K_{2,j}^{\epsilon,\mathcal{F}}$ such that

$$\mathbb{P}_{\mathcal{F}}\left\{\hat{a}_{jA+1}^{(j+1)A} \neq a_{jA+1}^{(j+1)A}\right\} \leq K_{1,j}^{\epsilon,\mathcal{F}}2^{-dC(\alpha_1-\epsilon)},$$

$$\mathbb{P}_{\mathcal{F}}\left\{\hat{b}_{jB+1}^{(j+1)B} \neq b_{jB+1}^{(j+1)B}\right\} \leq K_{2,j}^{\epsilon,\mathcal{F}}2^{-dC(\alpha_2-\epsilon)}.$$

In addition, $K_{1,j}^{\epsilon,\mathcal{F}}$ and $K_{2,j}^{\epsilon,\mathcal{F}}$ can not be too large in the following probabilistic sense,

$$\mathbf{P}\left(\left\{K_{1,j}^{\epsilon,\mathcal{F}} > K_1\right\} \cup \left\{K_{2,j}^{\epsilon,\mathcal{F}} > K_2\right\}\right) \leq \frac{1}{1-2^{-C\epsilon}}\left(\frac{K_{1,j}}{K_1^{\frac{\alpha}{\alpha-\epsilon}}} + \frac{K_{2,j}}{K_2^{\frac{\alpha}{\alpha-\epsilon}}}\right). \tag{60}$$

Proof. The proof is based on Markov's inequality and the union bound. The details can be found in (Oechtering and Rathi 2010). □

In the following theorem we prove a uniform bound on the constants $K_{1,j}^{\epsilon,\mathcal{F}}$ and $K_{2,j}^{\epsilon,\mathcal{F}}$ (over j).

Theorem 5.5. *Let $\{\mathcal{F}\}$ be the set of deterministic encoders with rate pair $R_1 = A/C$ and $R_2 = B/C$. The encoders are generated with probability measure $\mathbf{P}$ induced by the sequential encoding described in Subsection 5.2.2. For all $d > 0$ and all $j \geq 0$, let the average error probability (with respect to $\mathbf{P}$) for the two decoders is given by*

$$\mathbb{P}\left\{\hat{a}_{jA+1}^{(j+1)A} \neq a_{jA+1}^{(j+1)A}\right\} \leq K_1 2^{-dC\alpha_1}, \tag{61}$$

$$\mathbb{P}\left\{\hat{b}_{jB+1}^{(j+1)B} \neq b_{jB+1}^{(j+1)B}\right\} \leq K_2 2^{-dC\alpha_2}, \tag{62}$$

for some positive constants K_1 and K_2 which are independent of j and d. Then for every $\epsilon > 0$, $j \geq 0$, and $d > 0$, the error probability for almost every deterministic encoder $\mathcal{F}$ satisfies,

$$\mathbb{P}_{\mathcal{F}}\left\{\hat{a}_{jA+1}^{(j+1)A} \neq a_{jA+1}^{(j+1)A}\right\} \leq K_1^{\mathcal{F}} 2^{-dC(\alpha_1-\epsilon)}, \tag{63}$$

$$\mathbb{P}_{\mathcal{F}}\left\{\hat{b}_{jB+1}^{(j+1)B} \neq b_{jB+1}^{(j+1)B}\right\} \leq K_2^{\mathcal{F}} 2^{-dC(\alpha_2-\epsilon)}, \tag{64}$$

where $K_1^{\mathcal{F}}$ and $K_2^{\mathcal{F}}$ depends only on the encoder $\mathcal{F}$ and is independent of j and d.

Proof. The proof uses a generalization of Chebychev's inequality and arguments from (Sahai 2001a). The details can be found in (Oechtering and Rathi 2010). □

5.5 Conclusion

The bidirectional broadcast channel denotes a broadcast channel with two receivers where each receiver knows the message intended for the other. We consider streaming sources where messages of each user arrive as bit-stream to the encoder. We show that under maximum-likelihood decoding the average bit error probability decays exponentially in delay with positive exponent for all the rate pairs inside the capacity region. We also show the existence of deterministic codes which achieve exponentially decaying bit error probability with delay.

The bidirectional broadcast channel setup is the smallest cooperative communication network which benefits from the network coding idea. The study of this problem is motivated of from the problem of stabilizing two separated plants in an assembly line production setup. On the other hand it provides first insights for extending the anytime coding concept to larger communication networks. The study is based on information theoretic arguments in an idealized setting so that it provides principal insights and fundamental bounds for further studies.

6 Analysis of Anytime Transmission Scheme using UEP-LT Channel Codes

Networked control systems (NCS), i.e., distributed systems comprised of sensors, actuators and controllers communicating over a shared medium, is an area where control and communication theory intersect. In many control applications the system performance is typically very sensitive to delay in action. Furthermore, in real-time applications, the whole information may not be revealed prior to transmission. Therefore, in such applications transmission reliability over noisy communication networks cannot be ensured by coding an arbitrary long sequence due to delay constraints. In contrast, Sahai in (Sahai 2001b) proposed an information theoretical design methodology regarding channel coding for delay constrained systems.

In this section, we study the design of causal anytime channel codes for transmission over symmetric discrete memoryless channel in a cascaded source-channel coding system (Shirazinia, Bao and Skoglund 2011b). More specifically, we propose a transmission scheme, exploiting the unequal error protection (UEP) property, using Luby transform (LT) codes and sequential belief propagation (BP) decoding. Performance analysis is carried out where upper-bounds on the end-to-end distortion of a cascaded source-channel coding system are determined for both anytime repetition codes and anytime UEP-LT codes.

6.1 Background

The traditional approach to deal with channel imperfection such as packet loss, delay, data-rate limitations, and etc, when a dynamic source is considered, is dynamic programming (DP) (Teneketzis 2006). In many cases, the dynamic programming suffers from a serious problem which is called the curse of dimensionality, i.e., the computations become intractable when the number of states and decisions increases. Although approximate dynamic programming (ADP) (Powell 2007, Bao, Shirazinia and Skoglund 2010) can relax the problem to obtain some feasible solutions, it does not provide a unique solution to the original DP problem, and the computational complexity is still high in many cases.

In contrast with the traditional approach, Sahai in (Sahai 2001b) proposed the fundamental concept of anytime information transmission. There are two main features that distinguish traditional communication theory from the anytime concept. The anytime transmitter has only access to a part of the source message at

anytime, and the anytime receiver can use information in current channel output, as well as in previous channel outputs. In (Simsek 2004), bounds on anytime error exponent have been determined, and a time-sharing anytime channel code for a discrete memoryless channel (DMC) with perfect feedback is proposed. As a matter of fact, in many practical cases, a perfect feedback from the decoder to the remote transmitter is not possible because of diverse channel limitations. In (Como, Fagnani and Zampieri 2010), the distortion convergence rates for certain anytime coding schemes have been derived based on unequal error protection (UEP) property and assuming no channel feedback is available at the transmitter.

In the literature, rateless codes have been shown to have advantages in UEP applications (Rahnavard, Vellambi and Fekri 2007, Sejdinovic, Vukobratovic, Doufexi, Senk and Piechocki 2009). Rateless codes are a class of random sparse channel codes in which the encoder produces an unlimited number of symbols such that the decoder can recover the source symbols from a sufficiently large subset of channel outputs. Therefore, applying rateless codes is natural for encoding streaming data using anytime transmission schemes. Luby Transform (LT) codes (Luby 2002) are rateless error control codes suitable for erasure channels. From (Etesami and Shokrollahi 2006), we also know that LT and Raptor codes (Shokrollahi 2006) (concatenation of LT codes with high-rate LDPC outer encoder/decoder) perform very well on symmetric noisy channels such as binary symmetric channel (BSC) and binary-input additive white Gaussian noise (BI-AWGN) channel.

6.2 Problem Formulation

We shall assume that a scalar random variable $x_t \in \mathcal{X}$ is drawn according to a known distribution at time t where $\mathcal{X} \subseteq R$. Between the source and the destination there is a binary symmetric channel with bit cross-over probability ϵ. The inputs and outputs of the channel are denoted respectively by $y_t \in \{0, 1\}$ and $z_t \in \{0, 1\}$. The conditional probability of the channel is time-invariant, i.e., $\Pr(z_t|y_t) = \Pr(z|y)$.

In our anytime transmission scheme, the two main functional units, source coding and channel coding are considered separately. We employ truncated binary expansion as the source coding scheme to map x_t into a sequence of bits $(b_1, \ldots, b_{j_t})$, where j_t denotes the first j_t ($j_t \geq j_{t-1}$) bits available at the anytime encoder at time t. As will be clear later, the binary expansion fits well into the anytime framework that provides anytime reliability. In our problem formulation, we use the notation $\mathbf{x}_a^b = \{x_a, \ldots, x_b\}$ which denotes the evolution of a discrete-

time signal $x(t)$ from $t = a$ to $t = b$. The binary expansion is defined by a map $\mathcal{E}^s_{j_t} : \mathcal{X} \mapsto \mathcal{Y}^{j_t}$, where $\mathcal{Y}^{j_t}$ takes values from $\mathbf{b}_1^{j_t} = \{0, 1\}^{j_t}$. In fact, the choice of j_t depends on the rate of source coding, i.e., if the source generates a bit, b_j, every $1/R$ time units for a given source rate $R > 0$, then $j_t = \lceil Rt \rceil$. Without loss of generality, we assume that the source rate is $R=1$. The anytime channel encoder is described by a map $\mathcal{E}^c_t : \mathcal{Y}^{j_t} \mapsto \mathcal{Y}^t$ which outputs a bit at each time according to the function $E^c(\mathbf{b}_1^{j_t}) = y_t$.

At the remote receiver, the channel decoder, specified by a map $\mathcal{D}^c_t : \mathcal{Z}^t \mapsto \mathcal{Y}^{j_t}$, is allowed to exploit the information from the current, as well as the previous, received symbols, for the purpose of estimation, i.e., $\hat{\mathbf{b}}_1^{j_t} = D^c_t(\mathbf{z}_1^t)$. Finally, the source decoder can be written as the map $\mathcal{D}^s_{j_t} : \mathcal{Y}^{j_t} \mapsto \mathcal{X}$. The source decoder outputs the reconstructed value $\hat{x}_t$.

Herein, we consider a scalar linear plant with the system equation

$$x(t+1) = ax(t) + v(t) \quad , \quad t = 1, \ldots, T, \tag{65}$$

where $x(t) \in \mathcal{R}$ is the system state at time instant t, and $v(t) \in \mathcal{R}$ is the white process noise causes from inappropriate modeling. This open–loop system is unstable if $a > 1$ whose value is known at both the transmitter and the receiver. The system is triggered by the initial state $x(0)$ which has the uniform probability density function (pdf) on $[0, 1)$.

As in (Como et al. 2010), we consider that the variance of the process noise is much smaller than that of the initial state, hence the noise is negligible and the only source of uncertainty in the system is $x(0)$. Therefore, the system equation can be rewritten as $x(t) = a^t x(0)$, $t = 1, ..., T$. Following the notation of the anytime information pattern, at $t = 0$, the initial state is first quantized into j_t bits, and thereafter, a binary codeword, before it is sent over a binary symmetric channel. As a matter of fact, at each time instant t, a new channel coded bit representing the value of $x(0)$, referred to as $y_t \in \{0, 1\}$, is transmitted. At the receiver, the channel decoder obtains a noisy version of y_t, referred to as $z_t \in \{0, 1\}$, and starts providing estimates of the source value based on the current and previous received bits, i.e., $\mathbf{z}_1^t$. The reconstructed value $\hat{x}(0|t)$ is then given by applying the inverse binary expansion on the decoded bits. The block diagram of our dynamic system using an anytime code is sketched in Fig. 21.

We evaluate the system performance using end-to-end mean-squared error (MSE_{e2e}), defined as

$$\text{MSE}_{e2e}(t) \triangleq \mathbb{E}[(x(t) - \hat{x}(t))^2] = a^{2t}\mathbb{E}[(x(0) - \hat{x}(0|t))^2], \tag{66}$$

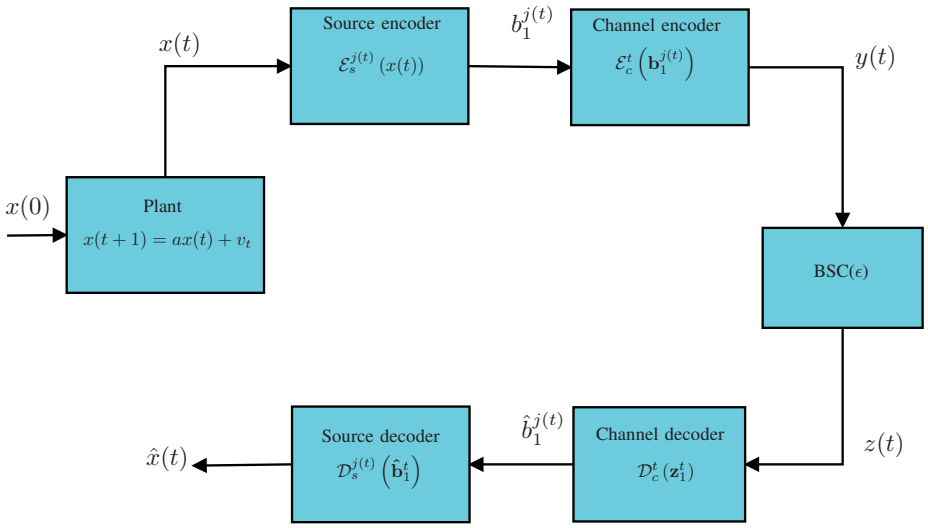


Figure 21: The functional diagram of a dynamic system using an anytime code.

where the last equality follows from $\hat{x}(t)=a^t\hat{x}(0|t)$, that is to say, in the absence of the process noise. The goal is to design a channel encoder/decoder pair in order to estimate $x(0)$ precisely out of the received data up to time t so that the MSE decays fast enough, i.e., $\lim_{t\to\infty}\text{MSE}_{e2e}(t)=0$. Thus, a^{2t} in (66) will be dominated by $\Delta_t^2 \triangleq \mathbb{E}[(x(0)-\hat{x}(0|t))^2]$ as time increases.

6.3 Anytime channel coding

In this subsection, we initialize with a motivating example inspired by repetition coding. Thereafter, we briefly present the proposed channel coding scheme.

6.3.1 Anytime UEP-repetition coding

As a part of the channel code, an anytime repetition code is used, which is a special block repetition code. More specifically, how frequently an information bit is coded is determined by its importance level. The first bit b_1 repeats most often since it is the most significant bit. For example, if the bit stream $\mathbf{b}_1^t$ is generated up to time t, the channel encoder might pick up a bit b_t according to the sequence $\{b_1; b_1b_2; b_1b_2b_3; \ldots\}$, where the position of the bit b_t is specified by the time instant t. The decoder receives $\{\hat{b}_1\hat{b}_1\hat{b}_2\ldots\}$, and employs majority logic decoding (MLD) so as to decide $\hat{b}_i{=}0$ $(1 \leq i \leq t)$ (or, 1) if it repeats more than $\hat{b}_i{=}1$ (or, 0) in the previous estimates. The decoder should also be able to decode if the number of received zeros or ones is equal. To handle this, it decides $\hat{b}_i{=}1$ or 0 using a Bernoulli trial.

The anytime repetition coding strategy, despite its relatively low computational complexity, can provide anytime reliability at the expense of zero rate and increasing delay with time. We bring up the idea of anytime rateless scheme which is suitable for applications of finite rate and delay.

6.3.2 Anytime UEP-LT encoding

In this section, we describe the encoding procedure which is an adaptation of the expanding window fountain (EWF) codes (Sejdinovic et al. 2009) fitting into the anytime transmission scheme. Let the information bits (input bits) denoted by b and the encoded bits (output bits) denoted by y at time $t=T$ be allocated to r overlapping information windows and r individual encoding windows. More specifically, at time t, the j^{th} $(j{=}1,\ldots,r)$ information window contains a portion

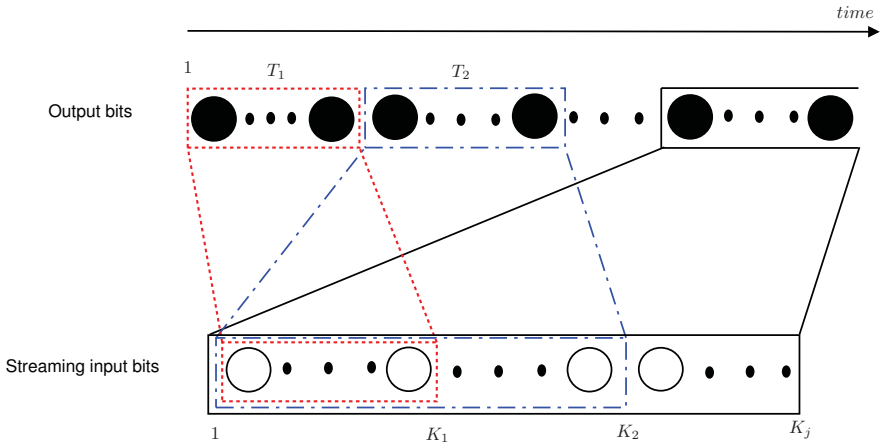


Figure 22: Anytime UEP-LT encoding scheme.

of the input bits $\mathbf{b}_1^t$ whose length will be referred to as K_j $(K_j > K_{j-1})$. Therefore, there are r levels of importance where each contains K_j–K_{j-1} bits.

Associated with the j^{th} information window, there exists a degree distribution $\Omega_j(x)$ according to which the bits in the corresponding window are chosen to be XOR'ed and generate an encoded bit y_t which is located in the j^{th} encoding window whose length is denoted by T_j. We emphasize here that the last information window does not necessarily contain all the information bits, $\mathbf{b}_1^T$, however, the encoding windows expand up to $t = T$, i.e., $\sum_{j=1}^{r} T_j = T$. It should be noted that K_j is a function of t as well, but for the ease of notation, we denote it by K_j. The encoding procedure is shown in Fig. 22.

Now, let $\Omega_m(x) = \sum_{d_m=1}^{D_m} \Omega_{d_m} x^{d_m}$ denote the output bit degree distribution corresponding to the $m^{th}(1 \leq m \leq j)$ encoding window, where Ω_{d_m} represents the probabilities of choosing the corresponding number of bits from the m^{th} information window. We define the output edge degree distribution as $\omega_m(x) = \sum_{d_m=1}^{D_m} \omega_{d_m} x^{d_m-1} = \Omega'_m(x)/\Omega'_m(1)$, where $\beta_m = \Omega'_m(1)$ is the average degree of an

output bit in the m^{th} encoding window. Similar degree distribution can be specified for the input bits as well. For this purpose, we analyze the input degree distribution of each level of importance individually. Let $\Lambda_m(x)=\sum_{n_m}\Lambda_{n_m}x^{n_m}$ be the input degree distribution of the $m^{th}(1\leq m\leq j)$ level of importance. The coefficients can be determined from the characteristics of output degree distribution in the following way. Let $I_m(x)=\sum_{n_m}I_{n_m}x^{n_m}$ denote the input degree distribution of the input bits in the m^{th} information window induced only by the edges connected to the output bits in the m^{th} encoding window. Therefore, $\Lambda_{n_m}=\prod_{i=m}^{j}I_{n_i}$, where only the input bits of the m^{th} level are considered. As shown in (Etesami and Shokrollahi 2006), $I_m(x)\approx e^{\alpha_m(x-1)}$ where $\alpha_m=I'_m(1)=\beta_m T_m/K_m$. Thus, input bit degree distribution of the m^{th} level of importance at $t\in T_j$ can be written as

$$\Lambda_m(x)=\exp\left[\left(\sum_{i=m}^{j-1}\alpha_i+\beta_j t/K_j\right)(x-1)\right]. \tag{67}$$

In a similar fashion, we introduce the input edge degree distribution of the m^{th} level as $\iota_m(x)=\sum_{n_m}\iota_{n_m}x^{n_m-1}$. It can be verified (see (Etesami and Shokrollahi 2006) for details) that $\iota_m(x)$ has asymptotically the same distribution as $\Lambda_m(x)$, therefore, $\Lambda'_m(1)=\iota'_m(1)=\sum_{i=m}^{j-1}\alpha_i+\beta_j t/K_j$.

6.3.3 Anytime UEP-LT decoding

Regarding the decoding procedure, the belief propagation (BP) algorithm is recognized as one of the successful algorithms in decoding graphical codes approaching Shannon capacity. It is an iterative process such that at each iteration input and output bits exchange messages, containing log-likelihood ratio (LLR). Let $B^{(l)}$ ($Y^{(l)}$) denote the messages passed from an input (output) bit b (y) to an output (input) bit at the l^{th} iteration of BP algorithm which are related as

$$\begin{aligned}\tanh(Y^{(l)}/2)&=\tanh(q_t/2)\prod_{adj(b)}\tanh(B^{(l)}/2)\\ B^{(l+1)}&=\sum_{adj(y)}Y^{(l)},\end{aligned} \tag{68}$$

where $q_t\triangleq\ln[\Pr(y_t=0|z_t)/\Pr(y_t=1|z_t)]$ denotes channel LLR at time t, and in the case of BSC, $q_t=\ln\left(\frac{1-\epsilon}{\epsilon}\right)(-1)^{z_t}$. Moreover, $adj(b)$ ($adj(y)$) represents the output (input) bits adjacent to the input (output) bit b (y). At the last iteration, the LLR of an input bit, which is obtained by $\sum_y Y^{(l)}$, is a measure to decide if the

decoded bit is 1 or 0, where the sum is over all output bits adjacent to the input bit. The anytime decoding sequentially applies the BP algorithm based on the assumption that the previous received bits $\mathbf{z}_1^{t-1}$ are available at the decoder.

To simplify the analysis of the BP algorithm, one can approximate the probability density of messages passed at each iteration by simple functions, instead of tracking the true densities. Commonly in practice, e.g., (Etesami and Shokrollahi 2006), $B^{(l)}$ and $Y^{(l)}$ are asymptotically approximated by Gaussian variables by which we only need to determine the mean and the variance.

6.4 Performance analysis

In this section, we propose the main results of this work regarding the anytime transmission schemes described in Section 6.3.

Theorem 6.1. *The source distortion at $t \in T_j$, subjected to the binary expansion of Section 6.2, is given by*

$$\Delta_s^2(t) = \frac{1}{3}\left(\frac{1}{4}\right)^{j_t}. \tag{69}$$

Theorem 6.2. *Given a BSC and the anytime UEP-LT codes of Section 6.3, the channel distortion $\Delta_{c,lt}^2(t)$ at $t \in T_j$ is upper bounded by*

$$\begin{aligned}&\Delta_{c,lt}^2(t) \leq \\ &\frac{1}{2}\sum_{i=1}^{j}\sum_{k=1+K_{i-1}}^{K_i} 2^{-2k}\exp\left(\sum_{m=i}^{j}\alpha_m(t)\left(\exp\left(-\mathbb{E}[Y_{m,i}^{(l)}]/4\right)-1\right)\right),\end{aligned} \tag{70}$$

where,

$$\begin{aligned}\alpha_m(t) &= \begin{cases}\beta_m T_m/K_m, & m = i,\cdots,j-1\\ \beta_j t/K_j, & m = j\end{cases}\\ \mathbb{E}\left[Y_{m,i}^{(l)}\right] &= \sum_{d_m}\omega_{d_m}\xi_{d_m}(\mu^{i,(l)})\\ \mu^{i,(l+1)} &= \iota_m'(1)\mathbb{E}[Y_{m,i}^{(l)}]\\ \xi_{d_m}(\mu^{i,(l)}) &= 2\mathbb{E}\left[\tanh^{-1}\left(\tanh(q/2)\prod_{m=1}^{d_m-1}\tanh(U_m/2)\right)\right].\end{aligned} \tag{71}$$

where U_m's ($m = 1, ..., d_m - 1$), representing the message $B^{(l)}$, are symmetric Gaussian distributed iid random variables, following $\mathcal{N}\left(\mu^{i,(l)}, 2\mu^{i,(l)}\right)$.

Corollary 6.3. *Given a BSC and the anytime UEP-LT code of Section 6.3, the end-to-end distortion at $t \in T_j$ is upper bounded by*

$$\Delta_{lt}(t) \leq \Delta_s(t) + \Delta_{c,lt}(t), \tag{72}$$

where $\Delta_s(t)$ and $\Delta_{c,lt}(t)$ are defined in (69) *and* (70)*, respectively.*

Theorem 6.4. *Given a BSC and the anytime repetition code of Section 6.3, the end-to-end distortion is upper bounded by*

$$\Delta_{re}(t) \leq \Delta_s(t) + \Delta_{c,re}(t), \tag{73}$$

where

$$\begin{aligned}
\Delta_s^2(t) &= \frac{1}{3}\left(\frac{1}{4}\right)^{j_t}, \\
\Delta_{c,re}^2(t) &= \sum_{\substack{k=1 \\ k:odd}}^{K} 2^{-2k} I_\epsilon\left(\frac{K-k+2}{2}, \frac{K-k+2}{2}\right) \\
&+ \sum_{\substack{k=1 \\ k:even}}^{K} 2^{-2k}\left[I_\epsilon\left(\frac{K-k+1}{2}, \frac{K-k+1}{2}\right)\right. \\
&\left.+\frac{1}{2}\binom{K-k+1}{\frac{K-k+1}{2}}\left(\epsilon-\epsilon^2\right)^{\frac{K-k+1}{2}}\right],
\end{aligned} \tag{74}$$

where $I_\epsilon(a,b)$ is the regularized incomplete beta function defined as

$$I_\epsilon(a,b) \triangleq \sum_{j=a}^{a+b-1} \binom{a+b-1}{j} \epsilon^j (1-\epsilon)^{a+b-1-j}.$$

The proofs are given in details in (Shirazinia et al. 2011b).

6.5 Numerical results

In this section, we quantify the performance of the anytime transmission schemes in terms of the end-to-end distortion and compare them with simulation results.

The number of information and encoding windows for the anytime UEP-LT codes is set to six windows with the lengths $\mathbf{K}$ = [10,12,30,40,50,60] and $\mathbf{T}$ = [60,65,180,250,270,300], respectively. Furthermore, the degree distributions are optimized for the BSC with cross-over probability $\epsilon = 0.11$ using the approach in (Shirazinia, Bao and Skoglund 2011a) which yields the degree distributions

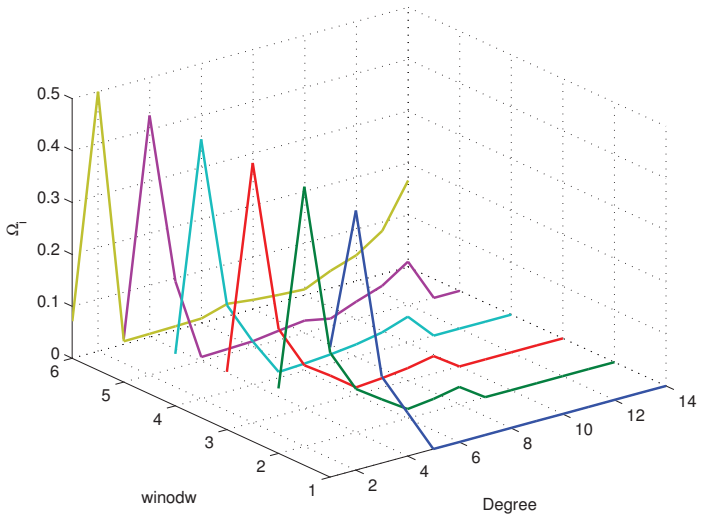


Figure 23: Degree distributions associated with the windows

shown in Fig. 23. The maximum order of the windows' degree distributions is set to $\mathbf{D}$=[4,8,9,10,12,14]. For the anytime UEP-LT codes, $\xi_{d_m}(\mu^i)$ in (71) is sampled using Monte-Carlo simulations by setting $\mu \in (0, 16]$ and the step-size as $\Delta\mu$=0.01. The number of BP iterations is also set to 50 rounds. The analytic and simulated results are reported in Fig. 24. It is worth pointing out that at lower time horizons, the asymptotic Gaussian assumption used in the proof of Theorem 6.2 does not hold. However, as time increases the upper bound provides the true bound on the distortion of the anytime UEP-LT scheme.

6.6 Conclusion

The problem of causal anytime channel codes of low complexity has been addressed in a source-channel coding system. In our studies, we have focused on a linear dynamic source and a binary symmetric channel. We have derived analytical upper bounds for the end-to-end distortion of two anytime transmission schemes. The first scheme was based on repetition coding exploiting the UEP encoding property and MLD decoding. Another anytime coding scheme was based on the UEP-LT encoding and sequential BP decoding with optimized degree dis-

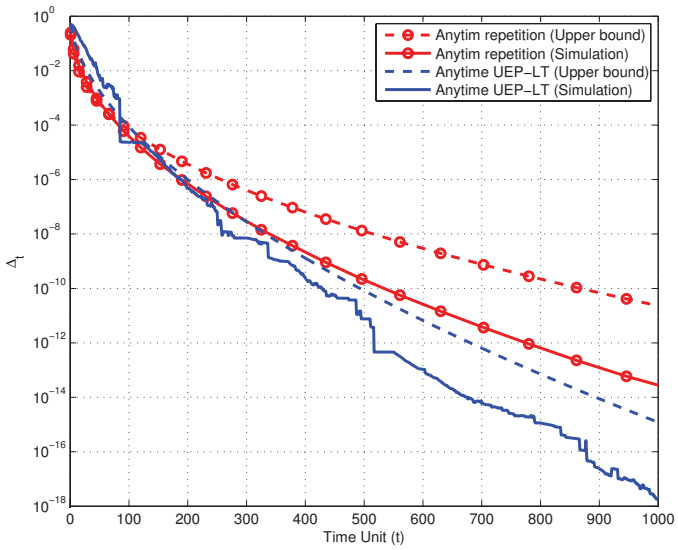


Figure 24: Comparison of simulated and analytic results of the anytime schemes

tribution. The upper-bounds have been compared with simulations in order to verify the accuracy of the analytical results.

The main advantage of this investigation is that they can give sufficient conditions for stability of an unstable closed-loop control system over a noisy communication channel and using a linear controller, which can be a subject of future work.

7 Conclusions

For Task 3.3 we have considered five problems in the scope of wireless transmission for control applications. Several fundamental problems have been studied and novel methods have been proposed.

- We derived a new transmission scheme for wireless communications over unknown non-orthogonal fading channel where several agents can transmit their data at the same time using the same frequency band. Two solutions have been derived. The first one acts by modulating the local data with doubly spread waveforms while the second one makes use of a new nonlinear precoder that allows viewing the received signal as a Volterra-like model. In both methods, the receiver has to deal with 3D data or signals admitting a trilinear model. The simulation results show that the proposed decoding methods provide good results.

- We considered a novel distributed source-and-channel coding scheme which can be used to reduce transmission rate or mitigate the effects of the channel noise in the case of analog transmission. From the Cramér-Rao lower bound, we observed general properties of analog distributed source-channel mappings. It was especially clear how the stretch factor influences the performance. From this observation we proposed two different mappings based on sinusoidal waveforms. The proposed transmission scheme was numerically evaluated and shown to perform well, particularly in the low-SNR regime. Since it requires no encoding or decoding delay, our proposed transmission scheme is suitable for delay-critical applications in wireless transmission for control.

- We optimized an instantaneous mapping for a relay node, which compresses received signals from different sources into one output symbol. This cooperative communication concept can be seen as an analog network coding strategy. From sum-rate simulations we see that the optimized mappings, in general, outperform linear mappings.

- We studied sequential coding for the bidirectional broadcast channel with streaming sources. Traditional information theoretic studies provide fundamental bounds on the achievable performance, but do not take processing delays into account. Rather, the new anytime coding concept seems to be the right approach to characterize the performance trade-off between reliability and delay suitable for control applications. Here, we extended the

concept to the bidirectional broadcast channel. Furthermore, we extended the existence proof of a deterministic code to such a multi-user channel.

- We have studied causal anytime channel codes of low complexity for communication over a binary symmetric channel in a source-channel coding system. We have derived analytical upper bounds for the end-to-end distortion of two anytime transmission schemes. The first scheme was based on repetition coding exploiting the UEP aspect and MLD decoding. Another anytime coding scheme was based on the UEP-LT encoding and sequential BP decoding with optimized degree distribution. The upper-bounds have been compared with simulations in order to verify the accuracy of the analytical results.

References

Ahlswede, R., Cai, N., Li, S.-Y. R. and Yeung, R. W.: 2000, Network Information Flow, *IEEE Transactions on Information Theory* **46**(4).

Bao, L., Shirazinia, A. and Skoglund, M.: 2010, Iterative encoder–controller design based on approximate dynamic programming, *IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC)*.

Caroll, J. and Chang, J.: 1970, Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition, *Psychometrika* **35**, 283–319.

Chang, C. and Sahai, A.: 2005a, Sequential random coding error exponents for degraded broadcast channels, *Allerton Conference on Control, Communications, and Computation*, Monticello, IL.

Chang, C. and Sahai, A.: 2005b, Sequential random coding error exponents for multiple access channels, *Proc. International Conference on Wireless Networks, Communications and Mobile Computing*, Vol. 2, pp. 1581 – 1586.

Chen, B. and Wornell, G. W.: 1998, Analog error-correcting codes based on chaotic dynamical systems, *IEEE Transactions on Communications* **46**(7), 881–890.

Chung, S.: 2000, *On the construction of some capacity-approaching coding schemes*, PhD thesis, MIT.

Como, G., Fagnani, F. and Zampieri, S.: 2010, Anytime reliable transmission of real-valued information through digital noisy channels, *SIAM Journal on Control Optimization* **48**(6), 3903 –3924.

Cover, T. M. and Gamal, A. A. E.: 1979, Capacity theorems for the relay channel, *IEEE Transactions on Information Theory* **25**(5), 572–584.

Cui, T., Ho, T. and Kliewer, J.: 2008, Memoryless relay strategies for two–way relay channels: performance analysis and optimization, *Proc. IEEE International Conference on Communications*.

de Almeida, A.-L.-F., Favier, G. and Mota, J.-C.-M.: 2007, PARAFAC-based unified tensor modeling for wireless communication systems with application to blind multiuser equalization, *Signal Processing* **87**(2), 337–351.

Draper, S. C., Chang, C. and Sahai, A.: 2005, Sequential Random Binning for Streaming Distributed Source Coding, *Proc. IEEE International Symposium on Information Theory*, Adelaide, Australia, pp. 1396–1400.

Etesami, O. and Shokrollahi, A.: 2006, Raptor codes on binary memoryless symmetric channels, *IEEE Trans. on Inf. The.* **52**(5), 2033 – 2051.

Farvardin, N. and Vaishampayan, V.: 1987, Optimal quantizer design for noisy channels: An approach to combined source–channel coding, *IEEE Transactions on Information Theory* **33**(6), 827–838.

Fernandes, C., Comon, P. and Favier, G.: 2010, Blind identification of MISO-FIR channels, *Signal Processing* **90**(2), 490–503.

Floor, P. A., Ramstad, T. A. and Wernersson, N.: 2007, Power constrained channel optimized vector quantizers used for bandwidth expansion, *International Symposium on Wireless Communication Systems*.

Forney, G. D.: 1974, Convolutional Codes II. Maximum-Likelihood Decoding, *Information and Control* **25**, 222–266.

Fuldseth, A. and Ramstad, T. A.: 1997, Bandwidth compression for continuous amplitude channels based on vector approximation to a continuous subset of the source signal space, *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Munich, Germany, pp. 3093–3096.

Gallager, R. G.: 1965, A Simple Derivation of the Coding Theorem and Some Applications, *IEEE Transactions on Information Theory* **18**(1), 3–18.

Gallager, R. G.: 1968, *Information Theory and Reliable Communication*, John Wiley & Sons, Inc., New York.

Gomadam, K. S. and Jafar, S. A.: 2006, On the capacity of memoryless relay networks, *Proc. IEEE International Conference on Communications*.

Harshman, R.: 1970, Foundation of the PARAFAC procedure: models and conditions for an "explanatory" multimodal factor analysis, *UCLA working papers in phonetics* **16**, 1–84.

Harshman, R.: 1972, Determination and proof of minimum uniqueness conditions for PARAFAC 1, *UCLA working papers in phonetics* **22**, 111–117.

Høst-Madsen, A. and Zhang, J.: 2006, Capacity bounds and power allocation for wireless relay channels, *IEEE Transactions on Information Theory* **51**(6), 2020–2040.

Karlsson, J. and Skoglund, M.: 2010, Optimized low-delay source–channel–relay mappings, *IEEE Transactions on Communications* (5), 1397–1404.

Katti, S., Maric, I., Goldsmith, A., Katabi, D. and Medard, M.: 2007, Joint relaying and network coding in wireless networks, *Proc. IEEE International Symposium on Information Theory*.

Khormuji, M. N. and Larsson, E. G.: 2008, Rate–optimized constellation rearrangement for the relay channel, *IEEE Communication Letters* **12**(9), 618–620.

Kibangou, A.: 2010, Distributed estimation over unknown fading channel, *2nd IFAC Workshop on Distributed Estimation and Control over Networked Systems (NECSYS 2010)*, Annecy, France.

Kibangou, A. and Favier, G.: 2009a, Blind equalization of nonlinear channels using tensor decompositions with code/space/time diversities, *Signal Processing* **89**(2), 133–143.

Kibangou, A. and Favier, G.: 2009b, Identification of parallel-cascade Wiener systems using joint diagonalization of third-order Volterra kernel slices, *IEEE Signal Processing Letters* **16**(3), 188–191.

Kibangou, A. and Favier, G.: 2009c, Non-iterative solution for PARAFAC with a Toeplitz matrix factor, *Proc. EUSIPCO*, Glasgow, Scotland.

Kim, S. J., Mitran, P. and Tarokh, V.: 2008, Performance Bounds for Bi-Directional Coded Cooperation Protocols, *IEEE Transactions on Information Theory* **54**(11), 5235–5241.

Kolda, T. and Bader, B.: 2009, Tensor decompositions and applications, *SIAM Review* **51**(3), 455–500.

Kramer, G., Gastpar, M. and Gupta, P.: 2005, Cooperative strategies and capacity theorems for relay networks, *IEEE Transactions on Information Theory* **51**(9), 3037–3063.

Kramer, G. and van Wijingaarden, A. J.: 2000, On the white Gaussian multiple-access relay channel, *Proc. IEEE International Symposium on Information Theory*.

Kruskal, J.: 1977, Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics, *Linear Algebra Applicat.* **18**, 95–138.

Kwon, H. and Cioffi, J.: 2008, Multi-user MISO broadcast channel with user-cooperating decoder, *Proc. of the IEEE 68 th Vehicular Technology Conference (VTC-Fall)*, Calgary, Canada, pp. 1–5.

Larsson, P., Johansson, N. and Sunell, K.-E.: 2005, Coded Bi-directional Relaying, *Proc. 5th Scandinavian Workshop on Ad Hoc Networks (ADHOC'05)*, Stockholm, Sweden.

Luby, M.: 2002, LT codes, *IEEE Symposium on Foundations of Computer Science* pp. 271 – 280.

Nion, D. and De Lathauwer, L.: 2008, An enhanced line search scheme for complex-valued tensor decompositions. Application in DS-CDMA, *Signal Processing* **88**(3), 749–755.

Oechtering, T. J. and Rathi, V.: 2010, Coding of streaming sources for the bidirectional broadcast channel, *5th International ICST Conference on Communications and Networking*. invited.

Oechtering, T. J., Schnurr, C., Bjelakovic, I. and Boche, H.: 2008, Broadcast Capacity Region of Two-Phase Bidirectional Relaying, *IEEE Transactions on Information Theory* **54**(1), 454–458.

Oechtering, T. J. and Skoglund, M.: 2010, Upper bound to error probability for coding on bidirectional broadcast channels, *Proc. International Conference on Telecommunications*, pp. 355 – 362.

Powell, W. B.: 2007, *Approximate Dynamic Programming: Solving the Curse of Dimensionality*, Wiley.

Rahnavard, N., Vellambi, B. and Fekri, F.: 2007, Rateless codes with unequal error protection property, *IEEE Trans. on Inf. The.* **53**(4), 1521 –1532.

Rajih, M., Comon, P. and Harshman, R.: 2008, Enhanced line search : a novel method to accelerate PARAFAC, **30**(3), 1148–1171.

Rankov, B. and Wittneben, A.: 2007, Spectral Efficient Protocols for Half-Duplex Fading Relay Channels, *IEEE Journal on Selected Areas in Communications* **25**(2), 379–389.

Sahai, A.: 2001a, *Any-Time Information Theory*, PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Sahai, A.: 2001b, *Anytime Information Theory*, PhD thesis, MIT.

Sahai, A. and Mitter, S. K.: 2006, The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link Part I: scalar systems, *IEEE Transactions on Information Theory* **52**(8), 3369–3395.

Sankaranarayanan, L., Kramer, G. and Mandayam, N. B.: 2004a, Capacity theorems for the multiple-access relay channel, *Proc. 42nd Annual Allerton Conference on Communication, Control and Computing*.

Sankaranarayanan, L., Kramer, G. and Mandayam, N. B.: 2004b, Hierarchical sensor networks: capacity theorems and cooperative strategies using the multiple-access relay channel model, *Proc. First IEEE Conference on Sensor and Ad Hoc Communications and Networks*.

Schetzen, M.: 1980, *The Volterra and Wiener theories of nonlinear systems*, John Wiley and Sons, Inc., New-York.

Sejdinovic, D., Vukobratovic, D., Doufexi, A., Senk, V. and Piechocki, R.: 2009, Expanding window Fountain codes for unequal error protection, *IEEE Trans. on Commun.* **57**(9), 2510 –2516.

Senol, H. and Tepedelenlioglu, C.: 2008, Performance of distributed estimation over unknown parallel fading channels, *IEEE Transactions on Signal Processing* **56**(12), 6057–6068.

Shirazinia, A., Bao, L. and Skoglund, M.: 2011a, Anytime source transmission using uep-lt channel coding, *European Wireless Conference*.

Shirazinia, A., Bao, L. and Skoglund, M.: 2011b, Distortion bounds on anytime source transmission using uep channel coding, *19th European Signal Processing Conference (EUSIPCO)*. to appear.

Shokrollahi, A.: 2006, Raptor codes, *IEEE Trans. on Inf. The.* **52**(6), 2551 –2567.

Sidiropoulos, N., Bro, R. and Giannakis, G.: 2000, Parallel factor analysis in sensor array processing, *IEEE Transactions on Signal Processing* **48**(8), 2377–2388.

Sidiropoulos, N. and Budampati, R.: 2002, Khatri-Rao space-time codes, *IEEE Transactions on Signal Processing* **50**(10), 2396–2407.

Sidiropoulos, N., Giannakis, G. and Bro, R.: 2000, Blind PARAFAC receivers for DS-CDMA systems, *IEEE Transactions on Signal Processing* **48**(3), 810–823.

Simsek, H. T.: 2004, *Anytime Channel Coding with Feedback*, PhD thesis, University of California, Berkely.

Slepian, D. and Wolf, J.: 1973, Noiseless coding of correlated information sources, *IEEE Transactions on Information Theory* **19**(4), 471–480.

Teneketzis, D.: 2006, On the structure of optimal real-time encoders and decoders in noisy communication, *IEEE Trans. on Inf. The.,* **52**(9), 4017–4035.

Tomasi, G. and Bro, R.: 2006, A comparison of algorithms for fitting the PARAFAC model, *Comp. Stat. Data Anal.* **50**(7), 1700–1734.

Trees, H. L. V.: 1968, *Detection, Estimation, and Modulation Theory. Part I*, Wiley.

Vaishampayan, V.: 1989, *Combined source–channel coding for bandlimited waveform channels*, PhD thesis, University of Maryland.

Vaishampayan, V. and Costa, S. I. R.: 2003, Curves on a sphere, shift-map dynamics, and error control for continuous alphabet sources, *IEEE Transactions on Information Theory* **49**(7), 1658–1672.

Wernersson, N., Karlsson, J. and Skoglund, M.: 2009, Distributed quantization over noisy channels, *IEEE Transactions on Communications* **57**(6), 1693–1700.

Wernersson, N., Skoglund, M. and Ramstad, T.: 2009, Polynomial based analog source–channel codes, *IEEE Transactions on Communications* **57**(9), 2600–2606.

Wimalajeewa, T. and Jayaweera, S.: 2007, Power efficient distributed estimation in a bandlimited wireless sensor network, *Proc. of 41st Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, USA.

Wong, T. and Lok, T.: 2000, Doubly spread DS-CDMA for efficient blind interference cancellation, *IEE Proc. Commun.* **147**(5), 299–304.

Wozencraft, J. M. and Jacobs, I. M.: 1965, *Principles of Communication Engineering*, Wiley.

Wu, Y., Chou, P. and Kung, S.-Y.: 2005, Information Exchange in Wireless Networks with Network Coding and Physical-Layer Broadcast, *Proc. of 39th Annual Conf. on Information Sciences and Systems*.

Wyner, A. D. and Ziv, J.: 1976, The rate-distortion function for source coding with side information at the decoder, *IEEE Transactions on Information Theory* **22**(1), 1–10.

Xie, L.-L.: 2007, Network Coding and Random Binning for Multi-User Channels, *Proc. 10th Canadian Workshop on Information Theory*, Edmonton, Alberta, Canada, pp. 85–88.

Yang, S. and Koetter, R.: 2007, Network coding over a noisy relay: a belief propagation approach, *Proc. IEEE International Symposium on Information Theory*.

Yao, S., Khormuji, M. N. and Skoglund, M.: 2008, Sawtooth relaying, *IEEE Communications Letters* **12**(9).

Yao, S. and Skoglund, M.: 2009, Analog network coding mappings in Gaussian multiple–access two–hop channels, *Proc. IEEE Information Theory Workshop*, Volvos.

Zaidi, A. A., Khormuji, M. N., Yao, S. and Skoglund, M.: 2009, Optimized analog network coding strategies for the white Gaussian multiple–access relay channel, *Proc. IEEE Information Theory Workshop*.

Ziehe, A., Laskov, P., Nolte, G. and Müller, K.-R.: 2004, A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation, *Journal of Machine Learning Research* **5**, 777–800.